

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12Q 1/68	A2	(11) International Publication Number: WO 00/66771 (43) International Publication Date: 9 November 2000 (09.11.00)
(21) International Application Number: PCT/EP00/03904 (22) International Filing Date: 30 April 2000 (30.04.00) (30) Priority Data: 60/131,984 30 April 1999 (30.04.99) US (71) Applicant (for all designated States except US): METHEXIS N.V. [BE/BE]; Onafhankelijkheidslaan 38, B-9000 Gent (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): ZABEAU, Marc [BE/BE]; Onafhankelijkheidslaan 38, B-9000 Gent (BE). STANSSENS, Patrick [BE/BE]; Constant Permekelaan 48, B-9830 Sint-Martens-Latem (BE). (74) Agent: DE CLERCQ, Ann; Ann De Clercq & Co B.V.B.A., E. Gevaertdreef 10 a, B-9830 Sint-Martens-Latem (BE).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: DIAGNOSTIC SEQUENCING BY A COMBINATION OF SPECIFIC CLEAVAGE AND MASS SPECTROMETRY		
(57) Abstract <p>The present invention is in the field of nucleic acid-based diagnostic assays. More particularly, it relates to methods useful for the "diagnostic sequencing" of regions of sample nucleic acids for which a prototypic or reference sequence is already available (also referred to as "re-sequencing"), or which may be determined using the methods described herein. This diagnostic technology is useful in areas that require such re-sequencing in a rapid and reliable way: (i) the identification of the various allelic sequences of a certain region/gene, (ii) the scoring of disease-associated mutations, (iii) the detection of somatic variations, (iv) studies in the field of molecular evolution, (v) the determination of the nucleic acid sequences of prokaryotic and eukaryotic genomes; (vi) identifying one or more nucleic acids in one or more biological samples; (vii) and determining the expression profile of genes in a biological sample and other areas.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 November 2000 (09.11.2000)

PCT

(10) International Publication Number
WO 00/66771 A3

- (51) International Patent Classification⁷: **C12Q 1/68**
- (21) International Application Number: PCT/EP00/03904
- (22) International Filing Date: 30 April 2000 (30.04.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/131,984 30 April 1999 (30.04.1999) US
- (71) Applicant (for all designated States except US):
METHEXIS N.V. [BE/BE]; Onafhankelijkheidslaan 38, B-9000 Gent (BE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **ZABEAU, Marc** [BE/BE]; Onafhankelijkheidslaan 38, B-9000 Gent (BE). **STANSSENS, Patrick** [BE/BE]; Constant Permekelaan 48, B-9830 Sint-Martens-Latem (BE).
- (74) Agent: **DE CLERCQ, Ann**; De Clercq, Brants & Partners cv, E. Gevaertdreef 10 a, B-9830 Sint-Martens-Latem (BE).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— With international search report.
- (88) Date of publication of the international search report:
8 February 2001
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: DIAGNOSTIC SEQUENCING BY A COMBINATION OF SPECIFIC CLEAVAGE AND MASS SPECTROMETRY

(57) Abstract: The present invention is in the field of nucleic acid-based diagnostic assays. More particularly, it relates to methods useful for the "diagnostic sequencing" of regions of sample nucleic acids for which a prototypic or reference sequence is already available (also referred to as "re-sequencing"), or which may be determined using the methods described herein. This diagnostic technology is useful in areas that require such re-sequencing in a rapid and reliable way: (i) the identification of the various allelic sequences of a certain region/gene, (ii) the scoring of disease-associated mutations, (iii) the detection of somatic variations, (iv) studies in the field of molecular evolution, (v) the determination of the nucleic acid sequences of prokaryotic and eukaryotic genomes; (vi) identifying one or more nucleic acids in one or more biological samples; (vii) and determining the expression profile of genes in a biological sample and other areas.

WO 00/66771 A3

INTERNATIONAL SEARCH REPORT

Int. Application No

PCT/EP 00/03904

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 20166 A (DEN BOOM DIRK VAN ; JURINKE CHRISTIAN (DE); HIGGINS G SCOTT (DE); L) 14 May 1998 (1998-05-14) cited in the application page 9; figure 6; example 4 page 22; figures 8,9; example 12 page 35; figures 20,21; example 20 page 41; example 21 page 62-63; example 27 page 79; claims	1-56
X	US 5 869 242 A (KAMB ALEXANDER) 9 February 1999 (1999-02-09) cited in the application the whole document	1-56
	-/-	

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

31 October 2000

Date of mailing of the international search report

07/11/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Reuter, U

INTERNATIONAL SEARCH REPORT

Int. Application No.

PCT/EP 00/03904

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 54571 A (INST MEDICAL W & E HALL ;FOOTE SIMON (AU); ELSO COLLEEN (AU); REID) 3 December 1998 (1998-12-03) cited in the application page 4-5 page 11-14; claims; examples 7,8 -----	1-5,17, 25-30, 42, 50-53,56
X	WO 97 33000 A (GENETRACE SYSTEMS INC) 12 September 1997 (1997-09-12) cited in the application page 10; claims; example 5 page 12, paragraph 2 page 13 page 16-17 page 28-29 page 33-40 -----	1-5,17, 20,21, 25-30, 42,45, 46, 50-53,56
A	US 5 643 798 A (BEAVIS RONALD C ET AL) 1 July 1997 (1997-07-01) column 1-2; claim 1 claims 16-18 -----	1-56
A	US 5 436 143 A (HYMAN EDWARD D) 25 July 1995 (1995-07-25) column 7-8 -----	1-50
A	WO 94 21663 A (MOLECULAR BIOLOGY RESOURCES) 29 September 1994 (1994-09-29) cited in the application page 1-15; examples 8,13 -----	1,17,18, 26,42,43

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l Application No

PCT/EP 00/03904



Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9820166 A	14-05-1998	US 5900481 A	04-05-1999
		US 6074823 A	13-06-2000
		US 6024925 A	15-02-2000
		US 6133436 A	17-10-2000
		AU 5106998 A	29-05-1998
		DE 19782095 T	23-03-2000
		EP 0954612 A	10-11-1999
		NO 992168 A	06-07-1999
		AU 5247298 A	29-05-1998
		DE 19782097 T	14-10-1999
		EP 0937097 A	25-08-1999
		NO 992167 A	05-07-1999
		WO 9820019 A	14-05-1998
		AU 5198098 A	29-05-1998
		DE 19782096 T	23-03-2000
		DE 29724251 U	17-08-2000
		DE 29724252 U	17-08-2000
		EP 0937096 A	25-08-1999
		NO 992169 A	06-07-1999
		WO 9820020 A	14-05-1998
US 5869242 A	09-02-1999	NONE	
WO 9854571 A	03-12-1998	AU 7514498 A	30-12-1998
		EP 0985148 A	15-03-2000
WO 9733000 A	12-09-1997	AU 2069597 A	22-09-1997
		CA 2248084 A	12-09-1997
		EP 0886681 A	30-12-1998
US 5643798 A	01-07-1997	US 5453247 A	26-09-1995
US 5436143 A	25-07-1995	AU 5873794 A	19-07-1994
		CA 2150670 A	07-07-1994
		EP 0675963 A	11-10-1995
		JP 8505053 T	04-06-1996
		WO 9414972 A	07-07-1994
		US 5516664 A	14-05-1996
		US 5629177 A	13-05-1997
		US 5514569 A	07-05-1996
		US 5602000 A	11-02-1997
WO 9421663 A	29-09-1994	US 5472872 A	05-12-1995
		AU 681650 B	04-09-1997
		AU 6524594 A	11-10-1994
		CA 2159081 A	29-09-1994
		EP 0690870 A	10-01-1996
		US 5604098 A	18-02-1997

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

14

Applicant's or agent's file reference METH-002-PCT		FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/EP00/03904	International filing date (day/month/year) 30/04/2000	Priority date (day/month/year) 30/04/1999
International Patent Classification (IPC) or national classification and IPC C12Q1/68		
Applicant METHEXIS N.V. et al.		
<p>1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.</p> <p>2. This REPORT consists of a total of 5 sheets, including this cover sheet.</p> <p><input checked="" type="checkbox"/> This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).</p> <p>These annexes consist of a total of 1 sheets.</p>		
<p>3. This report contains indications relating to the following items:</p> <ul style="list-style-type: none">I <input checked="" type="checkbox"/> Basis of the reportII <input type="checkbox"/> PriorityIII <input type="checkbox"/> Non-establishment of opinion with regard to novelty, inventive step and industrial applicabilityIV <input type="checkbox"/> Lack of unity of inventionV <input checked="" type="checkbox"/> Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statementVI <input type="checkbox"/> Certain documents citedVII <input type="checkbox"/> Certain defects in the international applicationVIII <input checked="" type="checkbox"/> Certain observations on the international application		
Date of submission of the demand 24/11/2000		Date of completion of this report 29.08.2001
Name and mailing address of the international preliminary examining authority:  European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465		Authorized officer Leber, T Telephone No. +49 89 2399 7195 

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/EP00/03904

I. Basis of the report

1. With regard to the **elements** of the international application (*Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17)*):

Description, pages:

1-63 as originally filed

Claims, No.:

1 as received on 06/08/2001 with letter of 06/08/0200

Drawings, sheets:

1/19-19/19 as originally filed

Sequence listing part of the description, pages:

1-8, as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☒ contained in the international application in written form.
- ☒ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/EP00/03904

- ☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes: Claims 1
	No: Claims
Inventive step (IS)	Yes: Claims 1
	No: Claims
Industrial applicability (IA)	Yes: Claims 1
	No: Claims

- 2. Citations and explanations**
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/EP00/03904

Re Item V

Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Basis for the assessment of novelty, inventive step and industrial applicability

- 1.1 This report takes into consideration the letter from the Applicant dated 06.08.2001.
- 1.2 The amendments filed with the letter of 06.08.2001 fulfill the requirements of Art 34(2)(b) PCT.
- 1.3 Reference is made to the following document/s/:

D1: WO 98 20166 A (DEN BOOM DIRK VAN ; JURINKE CHRISTIAN (DE); HIGGINS G SCOTT (DE); L) 14 May 1998 (1998-05-14) cited in the application

2. Novelty

- 2.1 Claim 1 appears to be novel (Art 33(2) PCT) as none of the documents cited in the ISR refers to a sequencing procedure whereby the target sequence is submitted to four separate base-specific cleavage reactions resulting in "non-ordered" fragments which means that, contrary to the methods disclosed in the prior art, the digestion is not carried out in a limited fashion but continued to completion.

3. Inventive step

- 3.1 Claim 1 differs from closest prior art document D1 in that non-ordered fragments are generated from the target nucleic acid for the purpose of nucleic acid sequencing instead of ordered fragments resulting from limited digestion reactions (e.g. D1, page 177, line 1- page 178, line 21; Fig. 77A-E). The technical problem

is to provide an improved method for sequencing nucleic acid molecules by MS. The solution referred to in claim 1 is to generate non-ordered fragments by guiding the digestion reaction to completion.

It appears that an inventive step (Art 33(3) PCT) can be acknowledged for this solution as none of the documents cited in the ISR indicates or disclose that nucleic acid can be sequenced on the basis of non-ordered fragments.

4. Industrial applicability

- 4.1 The subject-matter disclosed in the claims 1 of the present application appears to be industrially applicable (Art 33(4) PCT).

Re Item VIII

Certain observations on the international application

1. The expression "complementary cleavage reaction" in, for example, claim 1 lacks clarity (Art 6 PCT). A suitable definition should be included in the claim.
2. Claim 1 lacks clarity (Art 6 PCT) regarding whether or not the target nucleic acid is digested in four separate reactions each specific for a different base.
3. There are vague and imprecise statements in the description (page 21, lines 26-28; page 29, lines 2-5; page 34, lines 5-9; page 37, lines 1-2; page 43, lines 14-15; page 56, lines 3-7) of the present application implying that the subject-matter for which protection is sought may be different to that defined by the claims, thereby resulting in lack of clarity of the claims (Art. 6 PCT) when used to interpret them (Guidelines, Section IV, III-4.3a).

We claim:

1. A method for mass spectrometry based determination of the sequence of a target nucleic acid of unknown sequence present in a biological sample, said method comprising the steps of:
- 5 (a) deriving from said biological sample said target nucleic acid in a single stranded form;
- (b) subjecting said target nucleic acid obtained from step (a) to a set of four base-specific complementary cleavage reactions, wherein each cleavage
- 10 reaction generates a non-ordered set of fragments;
- (c) analyzing the sets of non-ordered fragments obtained from step (b) by mass spectrometry;
- (d) performing a systematic computational analysis on the mass spectra obtained from step (c) to assemble the sequence of said target nucleic acid;
- 15 and,
- (e) optionally, if the sequence is not uniquely defined after step (d), repeating steps (a) through (d), thereby generating modified forms of said target nucleic acid and/or different portions of said target nucleic acid, and performing supplementary mono- and/or di-nucleotide specific cleavage reactions rendering
- 20 supplementary sets of non-ordered fragments until the combined data converge into a unique sequence solution.

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference METH-002-PCT	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/EP 00/ 03904	International filing date (day/month/year) 30/04/2000	(Earliest) Priority Date (day/month/year) 30/04/1999
Applicant METHEXIS N.V		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

- b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

☒ contained in the international application in written form.

☒ filed together with the international application in computer readable form.

☐ furnished subsequently to this Authority in written form.

☐ furnished subsequently to this Authority in computer readable form.

☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☐ **Certain claims were found unsearchable** (See Box I).

3. ☐ **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No.

☐ as suggested by the applicant.

☐ because the applicant failed to suggest a figure.

☐ because this figure better characterizes the invention.

☒ None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 00/03904

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 20166 A (DEN BOOM DIRK VAN ; JURINKE CHRISTIAN (DE); HIGGINS G SCOTT (DE); L) 14 May 1998 (1998-05-14) cited in the application page 9; figure 6; example 4 page 22; figures 8,9; example 12 page 35; figures 20,21; example 20 page 41; example 21 page 62-63; example 27 page 79; claims ---	1-56
X	US 5 869 242 A (KAMB ALEXANDER) 9 February 1999 (1999-02-09) cited in the application the whole document --- -/--	1-56

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

31 October 2000

Date of mailing of the international search report

07/11/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Reuter, U

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/EP 00/03904

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 98 54571 A (INST MEDICAL W & E HALL ;FOOTE SIMON (AU); ELSO COLLEEN (AU); REID) 3 December 1998 (1998-12-03) cited in the application page 4-5 page 11-14; claims; examples 7,8 ---</p>	<p>1-5, 17, 25-30, 42, 50-53, 56</p>
X	<p>WO 97 33000 A (GENETRACE SYSTEMS INC) 12 September 1997 (1997-09-12) cited in the application page 10; claims; example 5 page 12, paragraph 2 page 13 page 16-17 page 28-29 page 33-40 ---</p>	<p>1-5, 17, 20, 21, 25-30, 42, 45, 46, 50-53, 56</p>
A	<p>US 5 643 798 A (BEAVIS RONALD C ET AL) 1 July 1997 (1997-07-01) column 1-2; claim 1 claims 16-18 ---</p>	<p>1-56</p>
A	<p>US 5 436 143 A (HYMAN EDWARD D) 25 July 1995 (1995-07-25) column 7-8 ---</p>	<p>1-50</p>
A	<p>WO 94 21663 A (MOLECULAR BIOLOGY RESOURCES) 29 September 1994 (1994-09-29) cited in the application page 1-15; examples 8, 13 -----</p>	<p>1, 17, 18, 26, 42, 43</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 00/03904

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9820166	A	14-05-1998	US 5900481 A	04-05-1999
			US 6074823 A	13-06-2000
			US 6024925 A	15-02-2000
			US 6133436 A	17-10-2000
			AU 5106998 A	29-05-1998
			DE 19782095 T	23-03-2000
			EP 0954612 A	10-11-1999
			NO 992168 A	06-07-1999
			AU 5247298 A	29-05-1998
			DE 19782097 T	14-10-1999
			EP 0937097 A	25-08-1999
			NO 992167 A	05-07-1999
			WO 9820019 A	14-05-1998
			AU 5198098 A	29-05-1998
			DE 19782096 T	23-03-2000
			DE 29724251 U	17-08-2000
			DE 29724252 U	17-08-2000
			EP 0937096 A	25-08-1999
			NO 992169 A	06-07-1999
			WO 9820020 A	14-05-1998
US 5869242	A	09-02-1999	NONE	
WO 9854571	A	03-12-1998	AU 7514498 A	30-12-1998
			EP 0985148 A	15-03-2000
WO 9733000	A	12-09-1997	AU 2069597 A	22-09-1997
			CA 2248084 A	12-09-1997
			EP 0886681 A	30-12-1998
US 5643798	A	01-07-1997	US 5453247 A	26-09-1995
US 5436143	A	25-07-1995	AU 5873794 A	19-07-1994
			CA 2150670 A	07-07-1994
			EP 0675963 A	11-10-1995
			JP 8505053 T	04-06-1996
			WO 9414972 A	07-07-1994
			US 5516664 A	14-05-1996
			US 5629177 A	13-05-1997
			US 5514569 A	07-05-1996
			US 5602000 A	11-02-1997
WO 9421663	A	29-09-1994	US 5472872 A	05-12-1995
			AU 681650 B	04-09-1997
			AU 6524594 A	11-10-1994
			CA 2159081 A	29-09-1994
			EP 0690870 A	10-01-1996
			US 5604098 A	18-02-1997

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G01N 33/483, C12Q 1/527, 1/68		A1	(11) International Publication Number: WO 98/54571 (43) International Publication Date: 3 December 1998 (03.12.98)
(21) International Application Number: PCT/AU98/00396 (22) International Filing Date: 28 May 1998 (28.05.98) (30) Priority Data: PO 7102 28 May 1997 (28.05.97) AU PO 7109 30 May 1997 (30.05.97) AU PP 1665 5 February 1998 (05.02.98) AU PP 3592 19 May 1998 (19.05.98) AU (71) Applicant (for all designated States except US): THE WALTER AND ELIZA HALL INSTITUTE OF MEDICAL RESEARCH [AU/AU]; Royal Parade, Parkville, VIC 3052 (AU). (72) Inventors; and (75) Inventors/Applicants (for US only): FOOTE, Simon [AU/AU]; 19 Park Street, Fairfield, VIC 3078 (AU). ELSON, Colleen [AU/AU]; 26 Christmas Street, Northcote, VIC 3070 (AU). SIMPSON, Richard [AU/AU]; 42 Stanley Street, Richmond, VIC 3121 (AU). REID, Gavin [AU/AU]; 330 Pascoe Vale Road, Essendon, VIC 3040 (AU). (74) Agents: HUGHES, E., John, L. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report.	
(54) Title: NUCLEIC ACID DIAGNOSTICS BASED ON MASS SPECTROMETRY OR MASS SEPARATION AND BASE SPECIFIC CLEAVAGE			
(57) Abstract A method of detecting a mutation or a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-ATOF MS and/or other equivalent procedure to produce a fingerprint of then oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NUCLEIC ACID DIAGNOSTICS BASED ON MASS SPECTROMETRY OR MASS SEPARATION AND BASE SPECIFIC CLEAVAGE**FIELD OF THE INVENTION**

5 The present invention relates generally to a method for detecting a mutation in a nucleic acid molecule. The method of the present invention does not require prior knowledge of a reference or wild-type nucleotide sequence nor does it require a gel electrophoresis step. The method of the present invention is particularly useful in identifying mutations and polymorphisms in genomic DNA and more particularly in the human genome and to determine and/or confirm the
10 nucleotide sequence of target nucleic acid molecules. The method of the present invention may also be automated.

BACKGROUND OF THE INVENTION

15 Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in a range of biotechnological fields. A particularly important area is the generation
20 of nucleotide mutants and the screening for and identification of such mutants. This in turn has implications, for example, in understanding the genetic basis behind certain disease conditions which is becoming of increasing relevance as the human genome is progressively sequenced.

An efficient and accurate method of mutation detection is crucial in implicating disease candidate
25 genes and in the screening programs which follow identification of disease causing mutations. Many human inherited and sporadic disorders are caused by small mutations including base substitutions, additions and deletions. Among these disorders are the Mendelian single gene disorders, sporadic somatic mutations causing cancers and complex genetic traits. Whilst some diseases are caused by a limited and well characterised set of mutations, most genetic diseases
30 are caused by one or more of a large range of mutations occurring anywhere within the gene. It is important, therefore, that a mutation detection protocol be able to scan a region of DNA,

- 2 -

identify any change and describe the resulting nucleotide differences from wild-type. With the increasing use of population molecular genetics and as clinicians begin to use mutation analysis as a *clinical* tool, there is a need to develop mutation detection protocols which can be automated, are less dependant on user expertise and are more accurate and reliable.

5

Current mutation detection protocols require either a gel based detection system or sequence specific primers. Gel based detection methods include direct sequencing of amplified DNA fragments and various techniques involving either cleavage of mismatched bases in heteroduplexes or mobility differences of single or partially denatured DNA strands.

10

Detection of mutations by DNA sequencing can provide good results in relation to accuracy and information about the position and nature of the mutation (Hattori *et al*, 1993), however, although advances have been made in this area, the technique is not fully automated and is labour intensive. Most mutations occur as heterozygotes and there are technical difficulties with the
15 ability of currently available computer software to identify two different nucleotide bases at a mutated residue.

Many mutation detection techniques exploit differential electrophoretic mobilities of DNA fragments with sequence differences. Single strand conformation polymorphism (SSCP) exploits
20 the fact that the secondary structure of a single strand of DNA is sequence based and, therefore, strands with even just one base difference will migrate at a different rate (Orita *et al*, 1989). This technique is again gel based and can lack sensitivity. Furthermore, the method cannot be readily automated and requires a large amount of labour due to the necessary gel step which in most cases must be optimised to the specific sample being analysed. They also do not give any
25 information about the position or nature of the change and do not routinely identify all mutations.

Mutation detection based on the identification of base pair mismatches in heteroduplex DNA strands is another method of identifying point changes. There are a number of techniques available that cleave DNA at mismatched base pairs in heteroduplex DNA. Mismatch cleavage
30 protocols include chemical and enzymatic mismatch cleavage. The techniques are also gel based. The chemical cleavage method uses osmium tetroxide to cleave at the mismatched base (Cotton

et al, 1988) followed by separation of cleaved products on denaturing gels. A major disadvantage of the chemical cleavage protocol is the use of extremely toxic chemicals.

Other methods for detection of known mutations include minisequencing allele specific
5 polymerase chain reaction (PCR), oligonucleotide probe arrays (Lipshutz *et al*, 1995) which requires knowledge of the sequence of wild-type and mutant. Although this technique is suitable for non-gel based detection methods, it is only useful for known mutations. Furthermore, the large number of oligonucleotides required to cover all known mutations in many genes makes this approach prohibitively expensive and labour intensive.

10

With the development of the matrix assisted laser desorption ionisation - time of flight mass spectrometer (MALDI-TOF MS), the ability to accurately determine the mass of biomolecules of a limited size has been achieved. Although detection of DNA fragments of up to 622 base pairs in length has been reported, large fragments cannot be accurately sized and a mass accuracy
15 of ± 3 bp is quoted (Liu *et al*, 1995). This level of accuracy is clearly insufficient for the detection and characterisation of base substitutions.

There is a need, therefore, to develop an effective and accurate means of detecting mutations in nucleic acid molecules. Preferably, the mutation detection system would be automatable.

20

In work leading up to the present invention the inventors developed a mutation detection system which exploits the accuracy of mass determination of MALDI-TOF MS and which is applicable for large DNA fragments. The method of the present invention does not require gel electrophoresis nor is prior knowledge of the nucleotide sequence necessary. The method of the
25 present invention is also capable of being automated.

SUMMARY OF THE INVENTION

Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined following the bibliography.

5

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

10

One aspect of the present invention contemplates a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide
15 fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

20

Another aspect of the present invention provides a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising amplifying said test nucleic acid molecule by polymerase chain reaction (PCR), subjecting the test amplified nucleic acid molecule to base specific cleavage to
25 generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a
30 difference of one or more nucleotides in said tested nucleic acid molecule.

Yet another aspect of the present invention is directed to a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising amplifying said test nucleic acid molecule by PCR, subjecting the test amplified nucleic acid molecule to base specific cleavage to generate
5 oligonucleotide fragments of from about 2 to about 1000 bases, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an
10 altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

Still yet another aspect of the present invention relates to a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic
15 acid molecule, said method comprising amplifying said test nucleic acid molecule and incorporating uracil residues, subjecting the test amplified nucleic acid molecule to uracil specific cleavage mediated by a uracil-N-glycosylase to generate oligonucleotide fragments of from about 2 to about 1000 bases, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the
20 oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

25 Another aspect of the present invention contemplates a computer programme capable of controlling a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or
30 other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an

altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

5 Yet another aspect of the present invention is directed to an apparatus capable of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said apparatus comprising means of subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent
10 procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

15

Still another aspect of the present invention provides a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide
20 fragments based on mass by MALDI-TOF MS and/or other equivalent procedure and subjecting said separated fragments to further separation means, such as post source decay (PSD) or other similar technique, to separate fragmentation products to generate a spectrum dependent on nucleotide sequence and then identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative
25 of a difference of one or more nucleotides in said tested nucleic acid molecule.

BRIEF DESCRIPTION OF THE FIGURES

- Figure 1** is a graphical representation showing mass spectrogram of cleavage products of two oligonucleotides, 1 and 2, which differ at two nucleotides, one produces a fragment with a 5 different nucleotide composition and the other introducing a new cleavage site. The two line thicknesses represent the overlaid tracings of the two different oligonucleotides. 1636.3 represents a thick line peak and 3190.9 represents a thin line peak. 1811.1 is a thin line peak and 1828.2 is a thick line peak. Kratos Kompact MALDI 4v51.2; % int. 100% = 24mV (thin); 81mV (thick).
- 10 **Figure 2** is a graphical representation showing mass spectrogram of reacted, separated products of normal TUB which represents a homozygote. Mode: linear; Accelerating Voltage: 20,000; Grid Voltage: 92.000%; Guide Wire Voltage 0-100%; Delay 125ON; Laser:1800; Scans Averaged: 128; Pressure: 9.94e-07; Low Mass Gate: 900.0; Negative Ions: ON.
- 15 **Figure 3** is graphical representation showing mass spectrogram of reacted, separated products of both TUB-M and TUB which represents a heterozygote. Mode: linear; Accelerating Voltage: 20,000; Grid Voltage: 92.000%; Guide Wire Voltage 0-100%; Delay 125ON; Laser:1800; Scans Averaged: 128; Pressure: 1.89e-06; Low Mass Gate 900.0; Negative Ions: ON.
- 20 **Figure 4** is a representation of the nucleotide sequence of IL-12 untranslated region PCR product used in Example 13. Primers are shown in bold. Expected cleavage products >2bp are underlined. The polymorphism is at position 97 and is indicated by asterisk. The polymorphism is a C to T change which results in a change of the cleavage products at that position from CGA to AGA in the forward strand and CAAGC to CAA in the reverse stand. The presence of C at position 97 results in a TaqI site and this allele is called "+", the other allele is respectively "-".
- 25 **Figure 5A** is a photographic representation of a TaqI restriction digest of IL-12 PCR products from +/- individuals (lanes 1, 4 and 5), a +/+ individual (lane 3) and a -/- individual (lane 2). The 30 124 bp fragment is cleaved by TaqI (where possible) to produce 97 and 27 bp fragments.

Figure 5B is a graphical representation showing linear MALDI-TOF spectra of cleavage products. The spectra on the left show a mass range of 1000 to 3500 and those on the right are the same spectra but show in detail the mass range from 1000 to 1700. Spectra *i* a and b are from a -/- individual, spectra *ii* a and b are from a +/+ individual and spectra *iii* a and b are from 5 a +/- individual. Observed masses are indicated above peaks. Arrows show the peaks that change between the two alleles.

Figure 6 is a graphical representation of the mass spectrum analysed using post source decay (PSD) on a MALDI-TOF instrument. Spectrum A is a 6mer of sequence CATCCT [SEQ ID NO:16] and spectrum B a 6mer of sequence CACCTT [SEQ ID NO:17]. Both have parent ion mass of 1727.2Da. Observed masses are shown above the peaks. PSD fragments are shown at an intensity magnification of five.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is predicated in part on a base specific cleavage reaction to generate a set of small oligonucleotides bounded by the base cleaved. The nucleic acid molecule may be
5 completely or only partially cleaved or digested. These fragments are then separated based on mass by MALDI-TOF MS. This generates a fingerprint of the nucleic acid fragment comprising a series of peaks where each peak represents the mass of each small cleavage product. As a result of the sensitivity of mass determination, each oligonucleotide of given length but different nucleotide composition produces a different mass. The mass of each peak, therefore,
10 corresponds to the nucleotide composition of the fragment as well as to its length. Consequently, any nucleotide substitution results in either a shifted peak due to the mass difference in the new cleavage fragment or, if the mutation changes the targeted base, a cleavage product containing a different number of bases.

15 Accordingly, one aspect of the present invention contemplates a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a
20 fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

25 Conveniently, screening is carried out by comparing the cleavage product masses of the reference or wild-type nucleic acid to those of the test sample. Mass changes corresponding to base changes are readily observed.

Accurate mass determination of these small fragments is possible allowing unambiguous
30 assignation of base composition of each oligonucleotide. This knowledge allows deduction of the nature of the mutation and, after specific cleavage at different bases and integration of the

- 10 -

data, the position of the mutation.

The method of the present invention is applicable to any nucleic acid molecule such as but not limited to DNA, genomic DNA, cDNA, plasmid DNA, satellite DNA, mRNA and other RNA
5 molecules as well as DNA:DNA, DNA:RNA and RNA:RNA hybrids. The present invention is particularly applicable to nucleic acid molecules amplified by, for example, polymerase chain reaction (PCR).

According to this aspect of the present invention, there is provided a method of detecting a
10 difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising amplifying said test nucleic acid molecule by polymerase chain reaction (PCR), subjecting the test amplified nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure
15 to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

20

A particularly preferred requirement is that the source of nucleic acid is cleavable to oligonucleotide fragments of from 2 bases to 1000 bases, preferably of from 3 bases to 500 bases, more preferably of from 4 bases to 100 bases and even more preferably of from 4 bases to 50 bases. Oligonucleotide fragments of from 4 bases to 40 bases are of particular usefulness
25 in practising the present invention.

Accordingly, the present invention is directed to a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising amplifying said test nucleic acid molecule by PCR, subjecting
30 the test amplified nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments of from about 2 to about 1000 bases, separating the resulting oligonucleotide

fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak
5 is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

The nucleic acid may be cleaved by a range of chemical molecules including enzymes. Enzymes are particularly preferred due to their specificity. One useful enzyme is uracil-N-glycosylase which cleaves DNA at uracil residues incorporated, for example, during a PCR. However, a
10 range of enzymes may be employed.

According to this embodiment, the present invention relates to a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising amplifying said test nucleic acid molecule and
15 incorporating uracil residues, subjecting the test amplified nucleic acid molecule to uracil specific cleavage mediated by a uracil-N-glycosylase to generate oligonucleotide fragments of from about 2 to about 1000 bases, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of
20 each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

The method of the present invention is predicated in part on the fact that any oligonucleotide
25 fragment differing in nucleotide composition between mutant and wild-type (or reference) sequences will be detected. The method has advantages over previously employed techniques and such advantages include the absence of a gel electrophoresis step thereby reducing time, expertise and need for separation equipment and the lack of dependance on toxic chemicals, such as osmium tetroxide. Whilst the present invention extends to the use of such chemicals in base
30 specific cleavage reactions, it is preferred to use an enzymatic reaction to cleavage the target nucleic acid molecule.

The method of the present invention is particularly useful in detecting previously unknown mutations. This is important as a screening mechanism for inherited diseases and cancers such as during pre-natal diagnosis, diagnosis of a suspected disease or screening for carriers of disease alleles. It also has applications in polymorphism analysis of populations and in studies of
5 evolution, drug resistance, virulence or attenuation of disease agents such as bacteria, viruses or protozoa.

The method may be carried out simultaneously or sequentially with an analysis of a reference to wild-type nucleic acid molecule. Both the test and reference nucleic acid molecules can then be
10 compared. Alternatively, the wild-type nucleic acid molecule may already have been analysed. Conveniently, this information may be stored electronically and upon completion of the analysis of the test nucleic acid molecule, both the test and reference sequences may then be compared manually, electronically or by a computer assisted means.

15 The method of the present invention may also be used to determine the nucleotide sequence of a nucleic acid molecule.

The nucleotide sequence may be completely determined or a partial sequence obtained for example, for selected nucleotides. The method of the present invention, therefore, permits the
20 rapid determination of a nucleotide sequence which will be invaluable, for example, in the efficient analysis of mutations.

The method of the present invention may be semi or fully automated and the present invention extends to apparatuses for automating the mutation detection assay. The apparatus may also be
25 electronically controlled by a computer programme to facilitate the automation and/or analysis process.

Accordingly, another aspect of the present invention contemplates a computer programme capable of controlling a method of detecting a difference of one or more nucleotides between a
30 nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide

fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same
5 procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

Yet another aspect of the present invention is directed to an apparatus capable of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a
10 reference nucleic acid molecule, said apparatus comprising means of subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak
15 relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

In a particularly preferred embodiment, the method of apparatus of the present invention also
20 employs a further fragment separation means such as but not limited to post source decay (PSD). PSD, for example, uses the dissociation of highly energised ions during their flight to the detector creating a second dimension. The ions are directed into an electric field of opposite polarity and are reflected. Smaller ions are reflected earlier and reach the detector first. As the spectrum from the decay is dependent on the nucleotide sequence of an oligonucleotide rather than the
25 nucleotide composition, this avoids missing mutations in an oligonucleotide having the same nucleotide composition as a reference oligonucleotide. Although PSD is one convenient fragment separation means, the present invention extends to other similar techniques to separate fragmentation products. Generally these techniques are based on mass although may also be based on electrophoretic mobility, base size, base charge, base pairing or other suitable criteria.

30

Accordingly, another aspect of the present invention provides a method of detecting a difference

- 14 -

of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure and subjecting
5 said separated fragments to further separation means to generate a spectrum dependent on nucleotide sequence and then identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

- 10 The MALDI-TOF MS analysis and further separation means may be done sequentially or simultaneously.

Preferably, the further separation means includes or comprises PSD or other similar techniques to separate fragmentation products.

15

The present invention is particularly useful in identifying and/or locating mutants in heterozygotes. Mutations are detectable on both strains or on one strand only.

- Yet another aspect of the present invention provides a method for identifying and/or locating a
20 mutation in one or more bases in a target nucleic acid molecule, subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a
25 reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

- Preferably, the separated fragments are subjected to further separation means such as but not
30 limited to PSD.

The present invention is further described by the following non-limiting Examples.

- 16 -

EXAMPLE 1

OLIGONUCLEOTIDES

Two test 22mers oligonucleotides with two bases different were used in this study

5

CCT CAT UTT TTU TTG TAA GAG G [SEQ ID NO:1]

CCT CGT UTT TTU TTG TUA GAG G [SEQ ID NO:2]

The different bases are shown in bold.

10

For the detection of point mutations (see Example 7), the following oligonucleotides are used:

TUB:

GGT GAC CTG AAC CAC CTC GTG CGT CCA GCC GTT CGT GGC TGT CCA GTC CGC

15 GAAC TCT GAC CTG CGC AAG [SEQ ID NO:3]

TUB-M:

GGT GAC CTG AAC CAC CTC GTG CGT CCA GCC GTT CGA GGC TGT CGA GTC

CGCGAA CTC TGA CCT GCG CAA G [SEQ ID NO:4]

TUB-F:

20 GGT GAC CTG AAC CAC CTC GT [SEQ ID NO:5]

TUB-R:

CTT GCG CAG GTC AGA GTT [SEQ ID NO:6]

TUB and TUB-M are used as template DNA and differ at three residues, bolded above, which
25 comprise two point mutations and one insertion (bracketed and bolded). TUB-F and TUB-R are
the "reverse" and "forward" primers used to amplify either TUB or TUB-M in a polymerase
chain reaction.

30

- 17 -

EXAMPLE 2

CLEAVAGE REACTION

The cleavage reactions were carried out using 100 pmol of oligonucleotide, 0.5 units uracil -N-
5 glycosylase (Perkin - Elmer) 1xPCR buffer (50mM KCl, 10mM Tris-HCl pH 8.3) (Perkin-Elmer)
in a 250 μ l reaction. The reaction mixture was incubated at 50°C for 20 minutes to allow
cleavage of the N-glycosidic bond at uracil. It was then heated for 15 minutes to 105°C to allow
degradation of the phosphate bonds at the basic sites. The mixture was then purified using anion
exchange resin to remove buffer salts and other impurities.

10

EXAMPLE 3

SAMPLE PURIFICATION

Qiagen Anion Exchange Resin was equilibrated in 5mM NH_4HCO_3 (Sigma) pH 8.4 (sodium
15 free). 40 μ l of the slurry was added to the reaction mixture and the DNA was allowed to bind
at room temperature for 5 minutes with gentle shaking. The beads were spun down in a bench
centrifuge and the supernatant discarded. The beads were then washed with 3x100 μ l volumes
of 5mM NH_4HCO_3 pH 8.4 (sodium free) with incubation and centrifugation between each wash.
The supernatant was discarded each time. The DNA fragments were then eluted using two 40 μ l
20 volumes of 0.5M NH_4HCO_3 pH 8.0 (sodium free), with incubation and centrifugation as before
but with the supernatant being kept. The supernatant was then evaporated to dryness on a
Savant Speedivac and resuspended twice in 20 μ l distilled water and evaporated to dryness to
remove any residual NH_4HCO_3 . The final product was resuspended in 5 μ l distilled water. The
final concentration being approximately 20pmol/ μ l.

25

EXAMPLE 4

THE POLYMERASE CHAIN REACTIONS AND DNA URACIL GLYCOSYLASE REACTION

30 20 μ l reactions were set up containing 2.5mM MgCl_2 , 2.5 mM dATP, dCTP, dGTP, 5 mM
dUTP, 0.5U Taq Gold (Perkin Elmer), 1.5 mM each TUB-F and TUB-R oligonucleotides and

- 18 -

2.4 fg of either TUB or TUB-M or a mix of both. PCR assays were incubated at 95°C for 15 minutes then cycled at 95°C - 15 seconds, 60°C - 35 seconds, 72°C - 35 seconds for 40 cycles. PCR reactions were pooled, each pool contained either 10 or 100 PCR reactions. Uracil DNA glycosylase (Perkin Elmer) was added at a ratio of 1U per 10 PCR reactions. Completeness of 5 digestion was confirmed by agarose gel electrophoresis.

EXAMPLE 5

PURIFICATION OF DIGESTED PCR PRODUCTS

10 Each DNA glycosylase reaction was loaded onto a C8 aquapore RP300 column equilibrated with 0.1M TEAA, the column washed with 0.1M TEAA at a flow rate of 0.5 ml/min and elute with 0.1M TEAA in 60% v/v CH₃CN. Peaks were collected. Column eluates were desiccated on a Savant Speedivac, evaporative centrifuge, resuspended in water to the original volume and redessicated. Pellets were resuspended in 5 ml H₂O. Mass spectrometric samples were prepared 15 as described in Example 6.

EXAMPLE 6

MASS ANALYSIS

20 3-Hydroxypicolinic acid is prepared at a concentration of 75mg/ml in 1:1 acetonitrile and water and stored at room temperature in a closed vial in the dark. A new matrix solution is prepared weekly. Cation exchange beads (Bio-Rad, 50W-X4, mesh size 100-200µm) in ammonium form were used to reduce interference from sodium and potassium adducts (Nordhoff *et al*, 1992). Samples were prepared as follows: 0.5µl matrix, 0.5µl sample (10pmol DNA) and 0.5µl cation 25 exchange resin were mixed on the slide and allowed to dry. The beads were then blown off with nitrogen gas. Samples were then analysed immediately.

Samples were run on the Kratos Kompact MALDI 4 with 337nm laser or a Perspective Voyager MALDITOF machine. Linear negative mode was used for all spectra. Fifty shots were fired at 30 power setting 70 to find a sweet spot and then a further 50 shots were fired at the sweet spot to obtain the spectrum.

- 19 -

EXAMPLE 7 SIMULATION

In order to assess the ability of this technique to detect mutations, a computer simulation was
5 designed. Two different stimulations were conducted, one that models a mutation occurring in
a haploid genome and the other modelling a mutation occurring in a diploid genome on the
background of a wildtype sequence.

In order to optimise the detection of mutations, four separate base specific cleavage reactions
10 have been performed using separated forward and reverse strands and two different base specific
reagents, in this case, thymidine and cytosine. A random library of exonic sequences has been
extracted from Genbank. This comprises 100,000 kb of coding sequence concatenated into one
file. Sequence strings of incremental length are removed from this file. A fingerprint for each
strand is generated. This is calculated by generating the sets of post cleavage fragments for each
15 base-specific reagent and sorting the non-redundant fragments. Mutant sequences are created
by mutating every residue in the wild-type sequence to each of three possible alternatives. The
fingerprint of each mutant is generated and compared to the wild-type fingerprints. If the
fingerprints are different, it is recorded as a successful detection and the next mutant examined.
If the first base specific cleavage reaction is unable to detect the mutation on the forward strand,
20 the reverse strand is tried and so on until the reverse strand of the second reagent fails. This
represents the total failure rate under the described conditions. Conceivably one could increase
the power of the technique by using all four base specific reagents on both strands.

EXAMPLE 8 DETECTION OF BASE MUTATIONS

25 Overlaid tracings from the mass spectrogram are presented in Figure 1. These show the cleavage
products of two oligonucleotides 1 and 2 [SEQ ID NO:1 and SEQ ID NO:2, respectively],
which differ at two nucleotides, one producing a fragment with a different nucleotide
30 composition and the other introducing a new cleavage site. The new fragments resulting from
these differences are easily separated. This example, observed masses deviate from calculated

- 20 -

by ± 0.02 -1%. This is sufficient to assign the correct base composition in this case, however, it is not sufficient to blindly assign base composition peaks from a sample of unknown sequence. A study has been done which concluded that all base compositions can be uniquely specified up to the 14mer level if one base has a known composition (ie. G=1 in the case of the study, or in our case, T=0) with a measurement of mass to within $\pm 0.01\%$. This is presently achievable, dependent on the mass analyser used and the sample quality and quantity (Pomerantz *et al.*, 1993).

Base specific cleavage and mass spectrometry is, therefore, able to differentiate between two identical length oligonucleotides with different nucleotide compositions and hence is able to differentiate between two sequences differing at one base (Table 1). Where a mutation changes the residue involved directly in the base specific cleavage reaction (a "U" residue in the case presented here), the difference in size of the resultant products is marked (Table 1). The accuracy of mass determination allows deduction of the base composition of each fragment and therefore, where the sequence is known, will enable deduction of the nature of the mutation.

Table 2 presents stimulation data for the haploid genome case and Table 3 presents the stimulation data where a mutation occurs in a diploid organism in the presence of a wild-type copy. These data are presented as cumulative "failure to identify" mutations based on both strands and two base specific cleavage reactions. Therefore, the last column, which is where the "C" reaction was unable to pick the mutation on the complementary strand represents the "total failure rate" of the technique under these conditions.

EXAMPLE 9

DETECTION OF POINT MUTATIONS

The method of the present invention has been employed on PCR products and is able to detect point mutations and an insertion in DNA that has been amplified using the polymerase chain reaction as discussed below. The PCR templates used, TUB and TUB-M are described in Example 1 and have three differences, two of which are point mutations and the third is an insertion/deletion. All of these differences are visible in the mass spectrograms (Figures 2 and

- 21 -

3). Figure 3 represents the reacted, separated products of both TUB-M and TUB. This is a reconstruction of a heterozygote. Figure 2 is reacted, separated products of TUB, representing, in this case a homozygote normal. Table 4 gives the expected masses for each fragment and the corresponding comments on whether they have been seen. All mutations were seen on either 5 both strands or on one strand only.

- 22 -

TABLE 1

oligo1:cleavage products		calc. mass	obs.mass
a	CCTCAT ¹	1810.2	1811.1
b	TTTT	1318.8	1318.4
c	TTGTAAGAGG ²	3190.0	3190.9

oligo2:cleavage products			
a	CCTCGT ³	1826.2	1828.2
b	TTTT	1318.8	1318.4
c	TTGT	1343.8	1343.5
d	AGAGG ⁴	1635.0	1636.3

¹ SEQ ID NO:18

² SEQ ID NO:19

³ SEQ ID NO:20

⁴ SEQ ID NO:21

Table 2.

Sequence length	Total number wildtype sequences	Total number mutated sequences	Number of "T" forward failures	% "T" forward failures	Number of "T" reverse failures	% "T" reverse failures	Number of "C" forward failures	% "C" forward failures	Number of "C" reverse failures	% "C" reverse failures
40	2500	300000	10569	3.52	97	0.03	1	0	0	0
60	1666	299880	14723	4.91	237	0.08	14	0	1	0
80	1250	300000	18908	6.30	401	0.13	26	0.01	0	0
100	1000	300000	22825	7.61	653	0.22	43	0.01	0	0
120	833	299880	26383	8.80	931	0.31	72	0.02	2	0
140	714	299880	29751	9.92	1225	0.41	93	0.03	8	0
160	625	300000	32644	10.88	1480	0.49	145	0.05	11	0
180	555	299700	35692	11.91	1955	0.65	188	0.06	21	0.01
200	500	300000	38848	12.95	2356	0.79	264	0.09	24	0.01
220	454	299640	40970	13.67	2828	0.94	348	0.12	32	0.01
240	416	299520	43995	14.69	3449	1.15	500	0.17	62	0.02
260	384	299520	46387	15.49	3923	1.31	541	0.18	88	0.03
280	357	299880	48443	16.15	4386	1.46	643	0.21	50	0.02
300	333	299700	50812	16.95	4904	1.64	759	0.25	103	0.03
320	312	299520	52651	17.58	5700	1.90	949	0.32	142	0.05
340	294	299880	54768	18.26	6130	2.04	1082	0.36	155	0.05
360	277	299160	56876	19.01	6621	2.21	1221	0.41	188	0.06
380	263	299820	59231	19.76	7445	2.48	1400	0.47	236	0.08
400	250	300000	60891	20.30	7906	2.64	1507	0.50	240	0.08
1000	100	300000	98902	32.97	27716	9.24	10255	3.42	3798	1.27

Table 3.

Sequence Length	Total number wildtype sequences	Total number mutated sequences	Number of "T" forward failures	% "T" forward failures	Number of "T" reverse failures	% "T" reverse failures	Number of "C" forward failures	% "C" forward failures	Number of "C" reverse failures	% "C" reverse failures
40	2500	300000	120699	40	39139	13	11468	4	3993	1
60	1666	299880	132516	44	48716	16	16455	5	6426	2
80	1250	300000	142523	48	57790	19	22043	7	9600	3
100	1000	300000	150441	50	65780	22	27368	9	12556	4
120	833	299880	156926	52	73381	24	32832	11	16347	5
140	714	299880	163973	55	81099	27	38470	13	20145	7
160	625	300000	169582	57	87388	29	43686	15	23767	8
180	555	299700	173319	58	92501	31	48016	16	27141	9
200	500	300000	178247	59	98957	33	53122	18	30973	10
220	454	299640	181728	61	103638	35	57622	19	34684	12
240	416	299520	184465	62	107959	36	62015	21	38330	13
260	384	299520	188025	63	112626	38	65898	22	41402	14
280	357	299880	191722	64	117075	39	70162	23	45504	15
300	333	299700	194210	65	120601	40	73900	25	48553	16
320	312	299520	196727	66	124257	41	77929	26	51963	17
340	294	299880	199290	66	127979	43	81308	27	54874	18
360	277	299160	200599	67	131116	44	84764	28	58213	19
380	263	299820	203752	68	134422	45	87956	29	60909	20
400	250	300000	205749	69	137686	46	91319	30	64056	21
420	238	299880	207767	69	140874	47	94747	32	67463	22
440	227	299640	209037	70	143182	48	96836	32	69713	23
460	217	299460	210530	70	145413	49	99414	33	71853	24
480	208	299520	212472	71	147958	49	102503	34	74722	25
500	200	300000	214285	71	149928	50	104741	35	77298	26
520	192	299520	215159	72	152227	51	107376	36	79507	27
540	185	299700	215902	72	153784	51	109511	37	81602	27

TABLE 4

EXPECTED TUB FRAGMENTS		FRAGMENTS NOT SEEN	
GGC	1045.6		
CCAC	1198.8		
CCACA [SEQ ID NO:22]	1512*		
CCAG	1318.8		
GGAC	1358.8		
CCAGCCG [SEQ ID NO:23]	2226.4		
GCGCAAG [SEQ ID NO:24]	2210.4		
GCGCAAGA [SEQ ID NO:25]	2523.6*		
CCGCGAAC [SEQ ID NO:26]	2539.6		
GGAGCACGCAGG [SEQ ID NO:7]	3880.4		
CGGCAAGCACCGACAGG [SEQ ID NO:8]	5374.4		
GGTGACCTGAACACCTCGTGCG [SEQ ID NO:9]	5888.8	PRIMER	
CAGGCGCTTGAGACTGGACGCGT [SEQ ID NO:10]	6258	PRIMER	
EXPECTED TUB-M FRAGMENTS			
CCAC	1198.8	END	
CGAG	1358.8		
GGAC	1358.8		
CGAGGC [SEQ ID NO:27]	1977.2		
CCAGCCG [SEQ ID NO:28]	2226.4		
GCGCAAG [SEQ ID NO:29]	2210.4		
CGACAGCC [SEQ ID NO:30]	2539.6		
CCGCGAAC [SEQ ID NO:31]	2539.6		
CGAACGGC [SEQ ID NO:32]	2579.6		
GGAGCACGCAGG [SEQ ID NO:11]	3880.4		
GGTGACCTGAACACCTCGTGCG [SEQ ID NO:12]	5888.8	PRIMER	
CAGGCGCTTGAGACTGGACGCGT [SEQ ID NO:13]	6258	PRIMER	

* Fragments obtained due to the terminal transferase activity of Taq polymerase which results in the addition of a dATP at the 3' end of the PCR product.

- 26 -

EXAMPLE 10

MODIFICATION DETECTION PROTOCOL

The method of Example 8 is employed except DNA polymerase enzymes are employed with the ability to incorporate both dNTPs and rNTPs. Specific cleavage reactions are performed on PCR products in which one of the nucleotides is substituted for rNTP. This permits the base specific cleavage reactions to be conducted in alkali at high temperature.

EXAMPLE 11

IDENTIFICATION OF MUTATION POSITION

The method of Example 8 employs Uracil-N-glycosylase which cleaves DNA at uracil. It is, therefore, a T reaction as uracil is replacing thymidine in the PCR product. In this Example, cleavage occurs at each of other bases so as to create sets of overlapping data to give information about the position of the mutation.

EXAMPLE 12

DETERMINATION OF NUCLEOTIDE SEQUENCE

The method of the present invention is used to determine a nucleotide sequence of a nucleic acid fragment. The method employed is substantially as described in Example 8.

EXAMPLE 13

DETECTION OF PREVIOUSLY UNKNOWN MUTATIONS

The method of the present invention is further demonstrated on a sequence polymorphism in the IL-12 gene. This previously unreported sequence change results in a TaqI RFLP and, therefore, can be followed by enzymatic digestion of PCR products.

Methods

Template DNA was genomic DNA from human volunteers of each possible genotype of the IL-12 polymorphism (ie. +/+, +/-, and -/-, where + is the presence of the Taq restriction site). PCRs were carried out in 20µl reactions in 192 well plates in a Corbett Thermocycler with the following reaction mixture: 50mM KCl, 10mM Tris-HCl pH 8.3, 25mM MgCl₂, 2.5mM dATP, dCTP and dGTP (Promega), 5mM dUTP (Boehringer Mannheim GmbH), 0.5U AmpliTaq Gold (Perkin Elmer), 0.4µM primers (Bresatec). After an initial 15min incubation at 95°C, the reactions were cycled 95°C 15 secs, 58°C 35sec, 72°C 35sec, for 40 cycles. 7 reactions were pooled for the homozygotes and 9 for the heterozygote. 1 unit of AmpErase Uracil-N-glycosylase (Perkin Elmer) was added to each pool and the reaction incubated at 50°C for 1 hour, followed by 30 minutes at 105°C. The extend of completion of the cleavage reaction was monitored by the absence of a band on an agarose gel. The cleavage reaction was monitored by the absence of a band on an agarose gel. The cleavage products were purified using reverse phase HPLC on a 100x2.1mm C8 aquapore RP300 column (Applied Biosystems). The flow rate was 0.5ml/min and absorbance was monitored at 254nm. The sample washed with 0.1M triethylaminoacetate (TEAA) and eluted in 0.1M TEAA/60% w/v acetonitrile and the fraction with absorbance at 254nm was collected and evaporated to dryness using a Savant Speedivac. The residue was resuspended in 100µl distilled deionised water and evaporated to dryness and then resuspended in 1µl water. 0.5µl of this was mixed with 0.5µl 3-hydroxypicolinic acid (saturated solution in 50% w/v acetonitrile and 0.5µl NH₄⁺ ion-exchange beads (BioRad, 50W-X4, mesh size 100-200µm) on a sample slide. The mass spectrometer used to characterise the reaction products was a Voyager BioSpectrometry Workstation from PerSeptive Biosystems. 128 laser pulses at power 1800 were averaged. Post Source Decay spectra were collected using a Kratos Kompact MALDI4 TOF mass spectrometer with 377nm laser and a curved field reflector in positive ion mode. Matrix and sample preparation as above. After scanning in linear mode for the sweet spot, the ion gate was set 34.8 Da above and 36.2 Da below the parent ion at 1727.2 Da. 200 profiles at 5 shots per profile were averaged. Spectra were corrected for the curved field.

Genotypes were confirmed by demonstrating the presence or absence of the TaqI restriction site

by digesting PCR products with TaqI restriction enzyme (Gibco-BRL) and analysing the products by agarose electrophoresis. DNA bands were stained with ethidium bromide.

A computer simulation of the method has been written and 100kb of random coding sequence from Genbank has been fed into it. The program takes discrete-length bites of sequence from a file of concatenated cDNA sequence from Genbank. Each base is mutated to each hypothetical variant of the original sequence by removing the cleaved base leaving the residual short strings. The mass spectrometry was modelled, fragments of different nucleotide composition being distinguishable and those of identical composition being indistinguishable. As quantitation is difficult on the MALDI, changes in peak height was not used as an indication of a change in underlying sequence. The program then compares "spectra" and tallies the number of mutations that were missed. The program can model the detection of a mutation in the presence of a wild-type sequence (heterozygote) or can model the differences between two homozygotes. In the first case a mutation can only be detected by the presence of a new peak and in the latter case, as well as the presence of a new peak, the disappearance of a peak can also signal a change. All four base specific cleavage reactions were used and reactions were performed on separated strands giving a total of 8 reactions per PCR product. Also the model has been refined to take account of the ability of post source decay (PSD) to identify changes in peaks containing a complex mix of oligonucleotides. In this case fragments of different sequence are distinguishable.

Results

A PCR assay was designed to incorporate the mutated region and then subjected to uracil -N glycosylase treatment. The products were purified and analysed by MALDI-TOF mass spectrometry. The sequence of the PCR primers and product along with the mutation are shown in Figure 4. The C to T change gives rise to a Taq RFLP and this can be seen in homozygote and heterozygote state in Figure 5. The spectra generated by the MALDI-TOF can also be seen in Figure 5. The expected and observed masses of the cleavage products from the two alleles are given in Table 5. The position of the mutation and deduction of the changed base is evident from study of this Table.

A limitation to the sensitivity of this method results from the lack of quantitative data available from the MALDI. When the fragment derived from the mutated sequence coincides with other fragments of identical nucleotide composition in the wild-type sequence, its disappearance will go undetected. Similarly, the appearance of a new fragment in the mutated sequence will go unnoticed if it has identical nucleotide composition to one or more other cleavage products. If both these conditions exist for all cleavage reactions, then the mutation will be missed. This technique, therefore, is not as advantageous for longer fragment as for small fragments.

To address this problem, the inventors employed a second dimension detection protocol on the MALDI-TOF machine. Post source decay (PSD) uses the dissociation of the highly energised ions during their flight to the detector as this second dimension. They are directed into an electric field of opposite polarity and are reflected. The smaller ions are reflected earlier and reach the detector first. As the spectrum from the decay is dependent on the sequence of the oligonucleotide (and not the nucleotide composition), the aforementioned limitation is bypassed, generating a method of mutation detection that is now extremely sensitive.

The utility of MALDI-TOF analysis with PSD is demonstrated in Figure 6 where two oligonucleotides of identical nucleotide composition are separated by MALDI-TOF using PSD. The resulting spectra are quite distinguishable. Sequence determination of small oligonucleotides is feasible using molecular dissociation methods and, therefore, the subject method extrapolates into an accurate resequencing protocol.

A computer simulation of data from the linear separation of cleavage products has been written. Using Genbank data, the expected number of base substitution that would be identified when comparing two homozygotes over a 250bp PCR distance is 98.5%. the comparable figure is 95% when a homozygote is compared to a heterozygote. If each mass peak from a base specific cleavage is analysed using a secondary dissociation technique, eg. PSD on the MALDI-TOF machine, then sensitivity of mutation detection improves dramatically. This has also been simulated and for a 1000bp fragment subjected to base specific cleavage, and analysed with PSD, 99% of all substitutions will be detected for a homozygote to heterozygote comparison and 99.8% when two homozygotes are compared.

- 30 -

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

BIBLIOGRAPHY

1. Cotton, R.G., Rodrigues, N.R., and Campbell, R.D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proceedings of the National Academy of Sciences of the United States of America* 85: 4397-401.
2. Hattori, M., Shibata, A., Yoshioka, K., and Sakaki, Y. (1993). Orphan peak analysis: a novel method for detection of point mutations using an automated fluorescence DNA sequencer. *Genomics* 15: 415-417.
3. Lipshultz, R.J., Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R., and Fodor, S.P. (1995). Using oligonucleotide probe arrays to access genetic diversity. *Biotech.* 19: 442-447.
4. Lui, Y-H., Bai, J., Zhu, X., Liang, X., Siemieniak, D., Venta, P.J., and Lubman, D.M. (1995). Rapid screening of genetic polymorphisms using buccal cell DNA with detection by Matrix - Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 9: 735-743.
5. Nordhoff, E., Ingendoh, A., Cramer, R., Overberg, A., Stahl, B., Karas, M., Hillenkamp, F., and Crain, P.F. (1992). Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared. *Rapid Commun. Mass Spectrom.* 6: 771-776.
6. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* 86: 2766-70.
7. Pomerantz, S.C., Kowalak, J.A. and McCloskey, J.A. (1993). Determination of oligonucleotide composition from mass spectrometrically measured molecular weight. *J. Am. Soc. Mass Spectrom.* 4: 204-209.

- 32 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: (Other than US): THE WALTER AND ELIZA HALL INSTITUTE OF
MEDICAL RESEARCH
(US only): FOOTE Simon, ELSO Colleen, SIMPSON Richard, REID
Gavin

(ii) TITLE OF INVENTION: A NOVEL METHOD

(iii) NUMBER OF SEQUENCES: 32

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: PCT INTERNATIONAL
(B) FILING DATE: 28 MAY 1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA

- (A) APPLICATION NUMBER: PO 7102
(B) FILING DATE: 28 MAY 1997
(A) APPLICATION NUMBER: PO 7109
(B) FILING DATE: 30 MAY 1997
(A) APPLICATION NUMBER: PP 1665
(B) FILING DATE: 5 FEBRUARY 1998
(A) APPLICATION NUMBER: PP3592
(B) FILING DATE: 19-MAY-1998

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: HUGHES, DR E JOHN L
(C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:

WO 98/54571

PCT/AU98/00396

- 33 -

- (A) TELEPHONE: +61 3 9254 2777
- (B) TELEFAX: +61 3 9254 2770
- (C) TELEX: AA 31787

- 34 -

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCATUTTT TUTTGTAAGA GG

22

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCTCGTUTTT TUTTGTAAGA GG

22

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGTGACCTGA ACCACCTCGT GCGTCCAGCC GTTCGTGGCT GTCCAGTCCG
CAAACCTCTGA CCTGCGCAAG

50

70

- 35 -

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGTGACCTGA ACCACCTCGTG CGTCCAGCCG TTCGAGGCTG TCGAGTCCGC
(G)AACTCTGAC CTGCGCAAG

50

69

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGTGACCTGA ACCACCTCGT

20

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTTGCGCAGG TCAGAGTT

18

- 36 -

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGAGCACGCAG G

11

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CGGCAAGCAC CGACAGG

17

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGTGACCTGA ACCACCTCGT GCG

23

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:

- 37 -

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CAGGCGCTTG AGACTGGACG CGT

23

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGAGCACGCA GG

12

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGTGACCTGA ACCACCTCGT GCG

23

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- 38 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CAGGCGCTTG AGACTGGACG CGT

23

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CACAACGGAA TAGACCCAAA AAGAUAAUUU CUAUCUGAUU UGCUUUAAAA
CGUUUUUUUA GGAUCACAAU GAUAUCUUUG CUGUAUUUGU AUAGUUCGAU
GCUAAAUGCU CAUUGAAACA AUCA

50

50

24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GUGUUGCCUU AUCUGGGUUU UUCUAUUAAA GAUAGACUAA ACGAAAUUUU
GCAAAAAAAU CCUAGUGUUA CUAUAGAAAC GACAUAAACA UAUCAAGCUA
CGATTTACGA GTAAC TTGT TAGT

50

50

24

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 base pairs
- (B) TYPE: nucleic acid

- 39 -

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CATCCT

6

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CACCTT

6

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCTCAT

6

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- 40 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTGTAAGAGG

10

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCTCGT

6

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGAGG

5

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 41 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCACA

5

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCAGCCG

7

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCGCAAG

7

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GCGCAAGA

8

- 42 -

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CCGCGAAC

8

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CGAGGC

6

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CCAGCCG

7

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 base pairs

- 43 -

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GCGCAAG

7

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CGACAGCC

8

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCGCGAAC

8

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

WO 98/54571

PCT/AU98/00396

- 44 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CGAACGGC

8

CLAIMS:

1. A method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.
2. A method according to claim 1 wherein the nucleic acid molecule to be tested is amplified by a polymerase chain reaction (PCR) prior to base specific cleavage.
3. A method according to claim 1 or 2 wherein the base specific cleavage results in oligonucleotide fragments of from about 2 bases to about 1000 bases.
4. A method according to claim 3 wherein the base specific cleavage results in oligonucleotide fragments of from about 3 bases to about 500 bases.
5. A method according to claim 4 wherein the base specific cleavage results in oligonucleotide fragments of from about 4 bases to about 100 bases.
6. A method according to any one of claims 1 to 5 wherein the base specific cleavage is uracil specific cleavage.
7. A method according to claim 6 wherein the uracil specific cleavage is mediated by uracil-N-glycosylase.
8. A method according to any one of claims 1 to 7 further comprising subjecting

fragmentation products to further separation (PSD) to generate a spectrum from decay dependent on the nucleotide sequence of the oligonucleotide.

9. A method according to claim 8 wherein the further separation of fragmentation products is by post source decay (PSD).

10. A computer programme capable of controlling a method of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said method comprising subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

11. A method according to claim 9 wherein the nucleic acid to be tested is amplified by PCR prior to base specific cleavage.

12. A method according to claim 9 or 10 wherein the base specific cleavage results in oligonucleotide fragments of from about 2 bases to about 1000 bases.

13. A method according to claim 9 wherein the base specific cleavage results in oligonucleotide fragments of from about 3 bases to about 500 bases.

14. A method according to claim 10 wherein the base specific cleavage results in oligonucleotide fragments of from about 4 bases to about 100 bases.

15. A method according to any one of claims 9 to 13 wherein the base specific cleavage is uracil specific cleavage.

16. A method according to claim 14 wherein the uracil specific cleavage is mediated by uracil-N-glycosylase.

17. A method according to any one of claims 10 to 16 further comprising the further separation of fragmentation products to generate a spectrum from decay dependent on the nucleotide sequence of the oligonucleotide.

18. A method according to claim 17 wherein the further separation of fragmentation products is by post source decay (PSD).

19. An apparatus capable of detecting a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule, said apparatus comprising means of subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

20. An apparatus according to claim 19 further comprising further fragmentation separation means to generate a spectrum from decay dependent on the nucleotide sequence of the oligonucleotide.

21. An apparatus according to claim 20 wherein the further fragmentation separation means is post source decay (PSD).

22. Use of MALDI-TOF in the detection of a difference of one or more nucleotides between a nucleic acid molecule to be tested and a reference nucleic acid molecule.

23. Use according to claim 22 further comprising use of PSD to generate a spectrum for

decay dependent on the sequence of an oligonucleotide.

24. A method for identifying and/or locating a mutation in one or more bases in a target nucleic acid molecule, subjecting the test nucleic acid molecule to base specific cleavage to generate oligonucleotide fragments, separating the resulting oligonucleotide fragments based on mass by MALDI-TOF MS and/or other equivalent procedure to produce a fingerprint of the oligonucleotide fragments comprising one or more peaks wherein a peak represents the mass of each fragment and identifying an altered peak relative to a reference nucleic acid molecule subjected to the same procedure wherein the presence of an altered peak is indicative of a difference of one or more nucleotides in said tested nucleic acid molecule.

25. A method according to claim 24 wherein the nucleic acid molecule to be tested is amplified by a polymerase chain reaction (PCR) prior to base specific cleavage.

26. A method according to claim 24 or 25 wherein the base specific cleavage results in oligonucleotide fragments of from about 2 bases to about 1000 bases.

27. A method according to claim 26 wherein the base specific cleavage results in oligonucleotide fragments of from about 3 bases to about 500 bases.

28. A method according to claim 27 wherein the base specific cleavage results in oligonucleotide fragments of from about 4 bases to about 100 bases.

29. A method according to any one of claims 24 to 28 wherein the base specific cleavage is uracil specific cleavage.

30. A method according to claim 29 wherein the uracil specific cleavage is mediated by uracil-N-glycosylase.

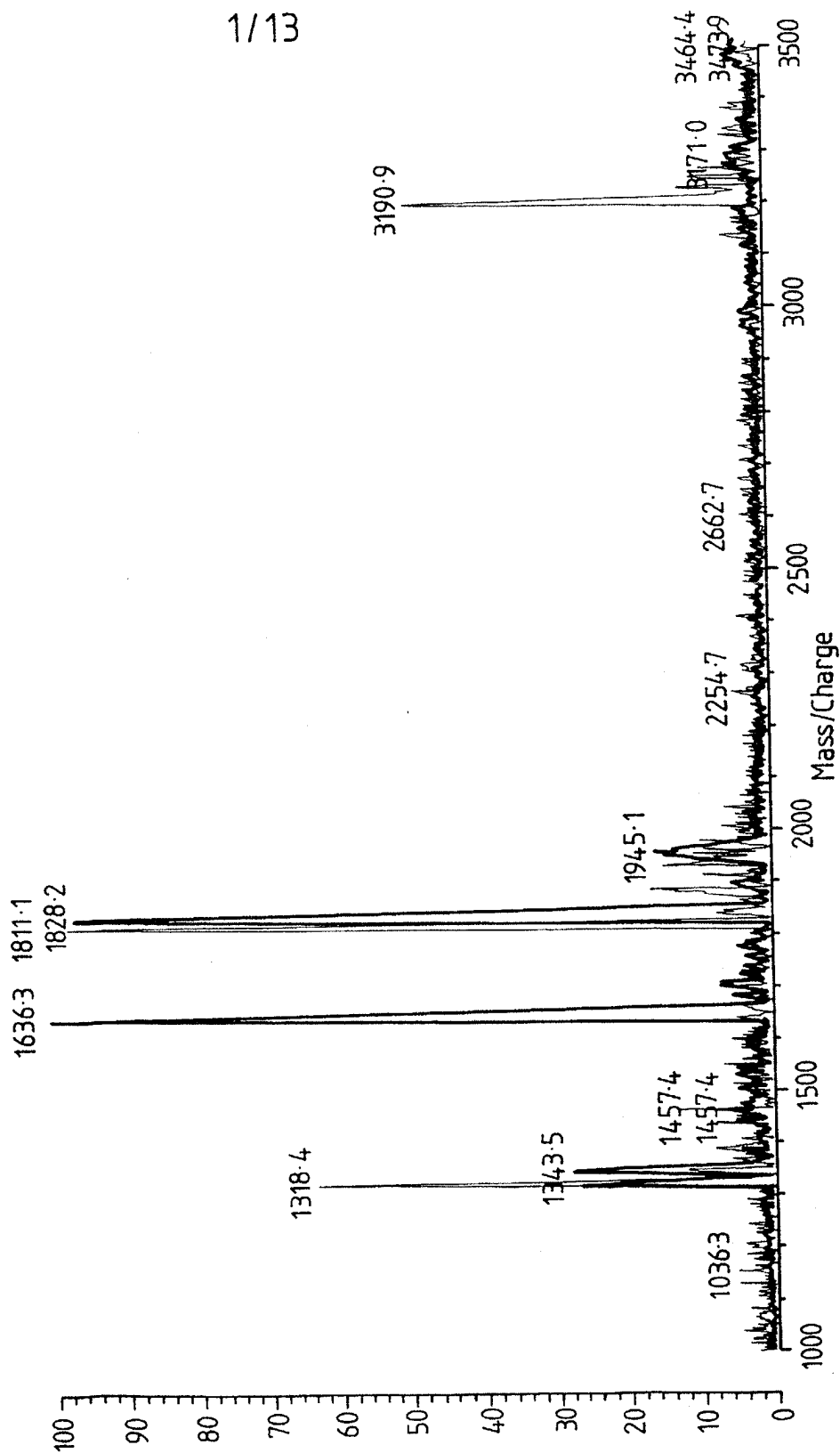
31. A method according to any one of claims 24 to 30 further comprising subjecting fragmentation products to further separation (PSD) to generate a spectrum from decay

dependent on the nucleotide sequence of the oligonucleotide.

32. A method according to claim 31 wherein the further separation of fragmentation products is by post source decay (PSD).

1/13

FIG 1



SUBSTITUTE SHEET (RULE 26)

2/13

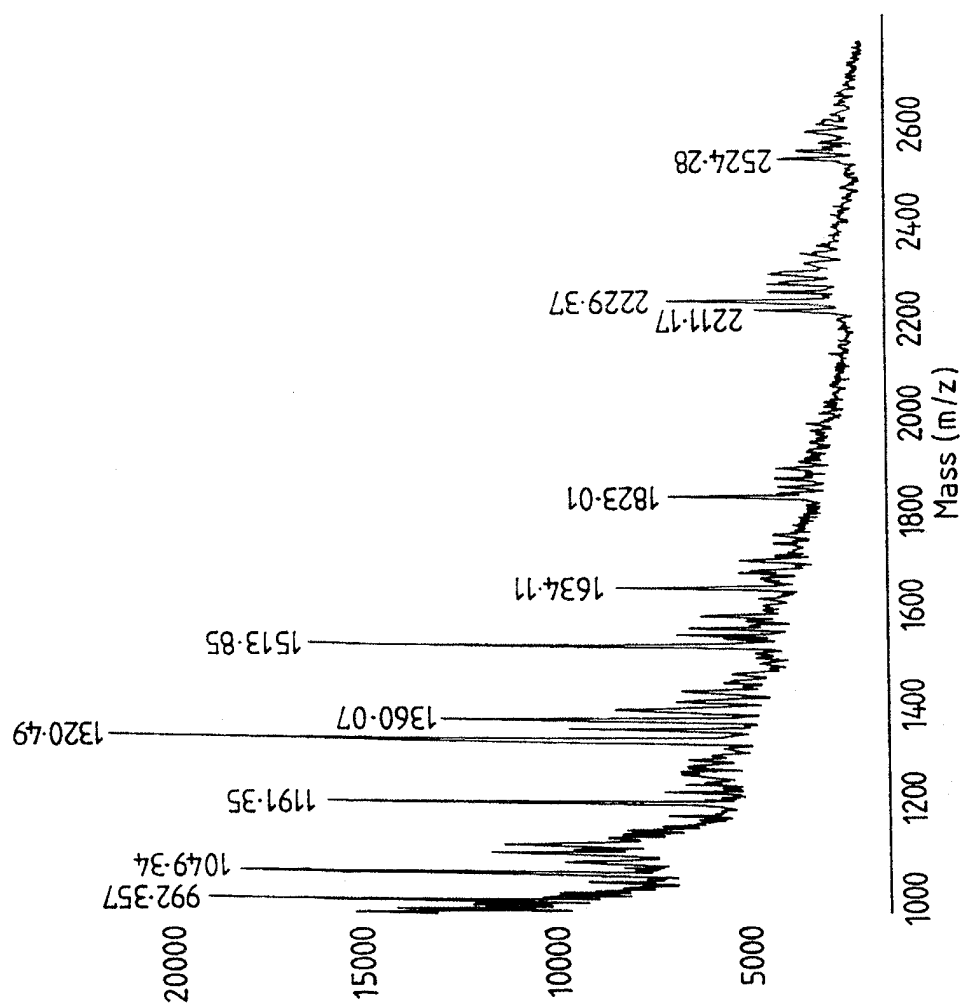
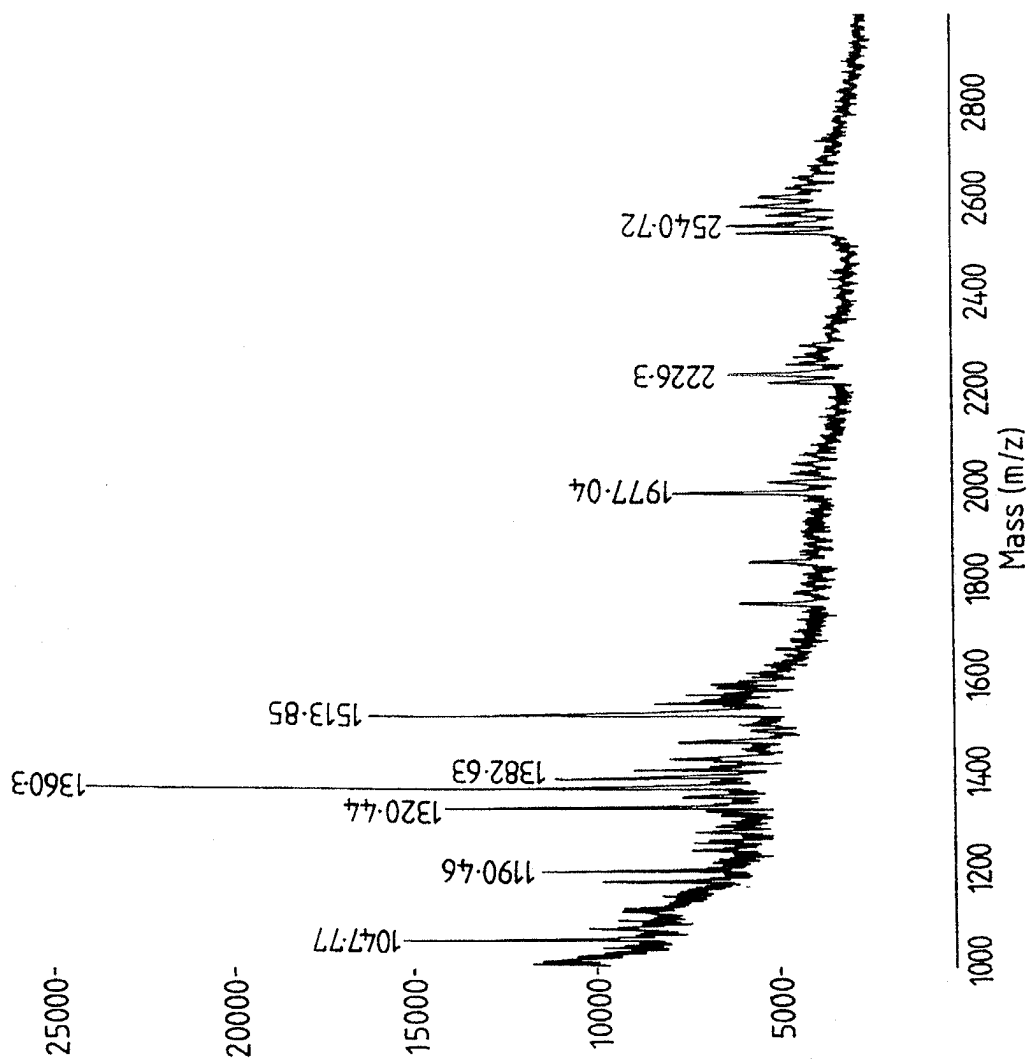


FIG 2

SUBSTITUTE SHEET (RULE 26)

3/13

**FIG 3**

SUBSTITUTE SHEET (RULE 26)

4/13

CAC AAC GGA ATA GAC CCA AAA AGA UAA UUU CUA UCU
GUG UUG CCU UAU CUG GGU UUU UCU AUU AAA GAU AGA

GAU UUG CUU UAA AAC GUU UUU UUA GGA UCA CAA UGA
CUA AAC GAA AUU UUG CAA AAA AAU CCU AGU GUU ACU

UAU CUU UGC UGU AUU UGU AUA GUU CGA UGC UAA AUG
AUA GAA ACG ACA UAA ACA UAU CAA GCU ACG ATT TAC *

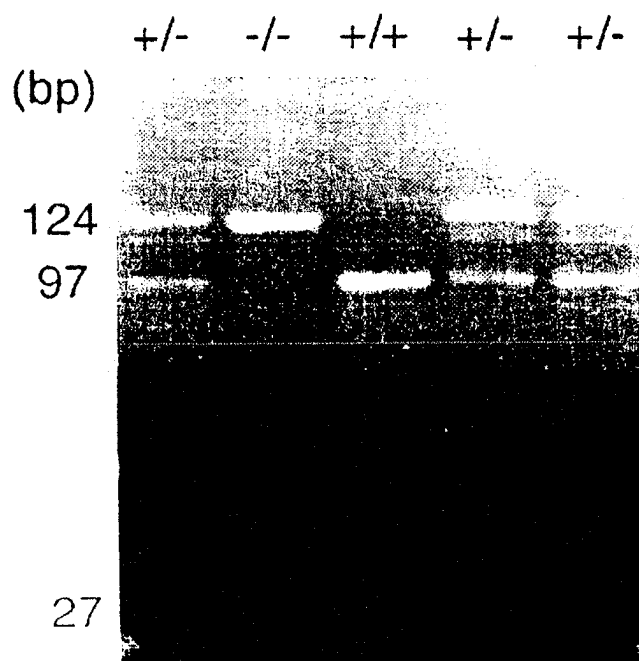
CUC AUU GAA ACA AUC A
GAG TAA CTT TGT TAG T

FIG 4

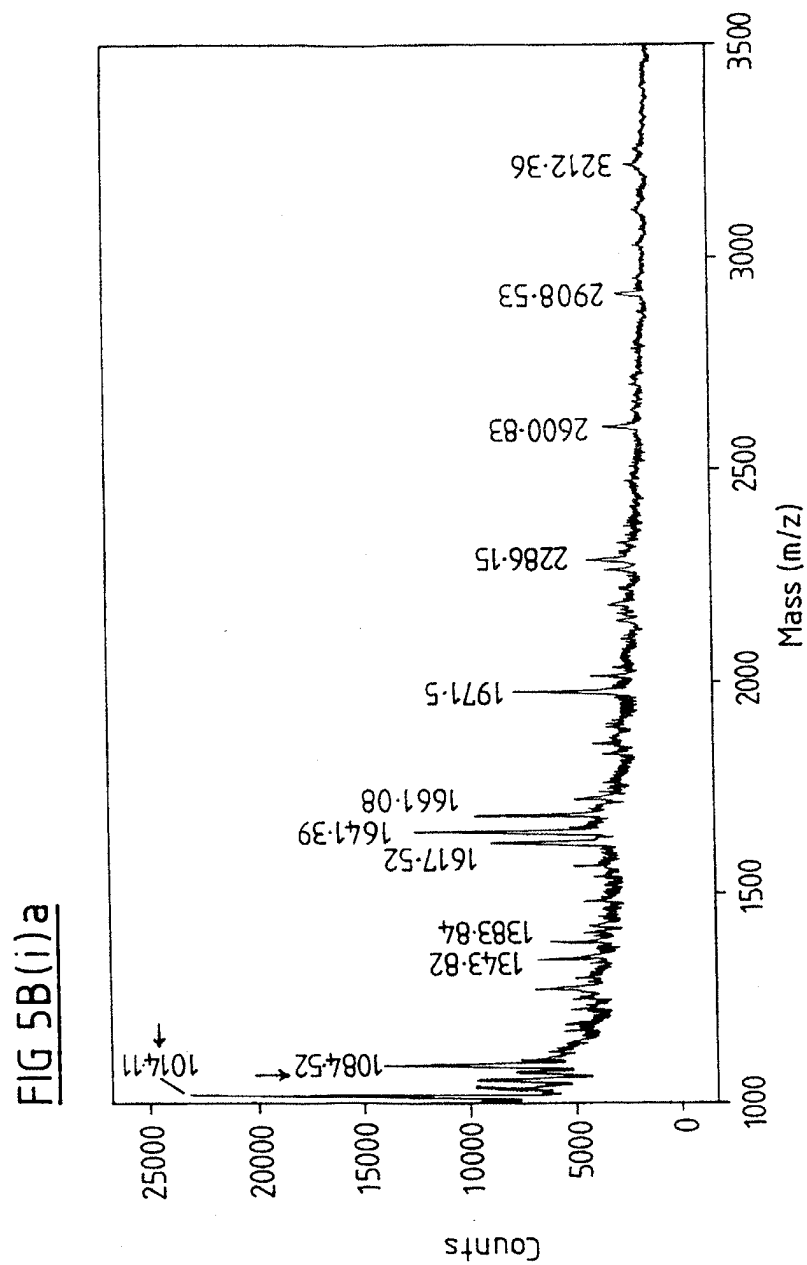
5/13

Figure 5A

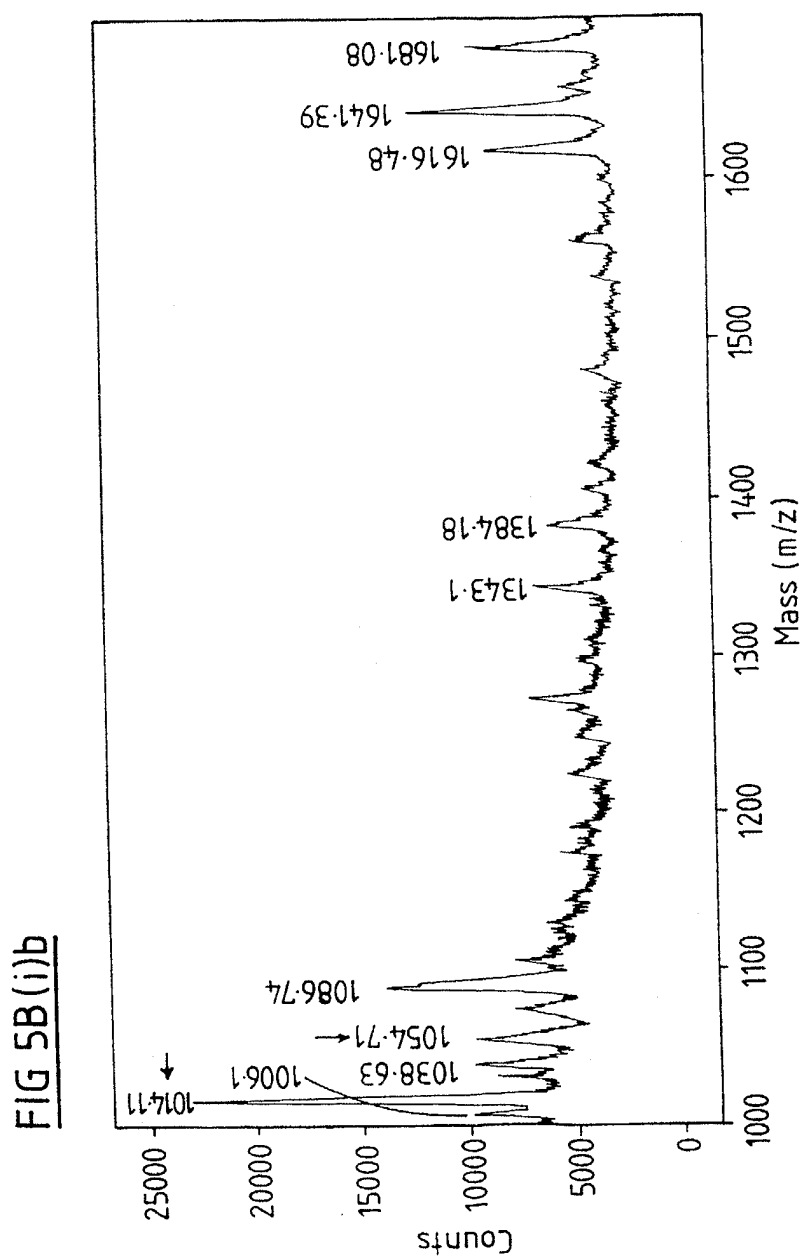
A



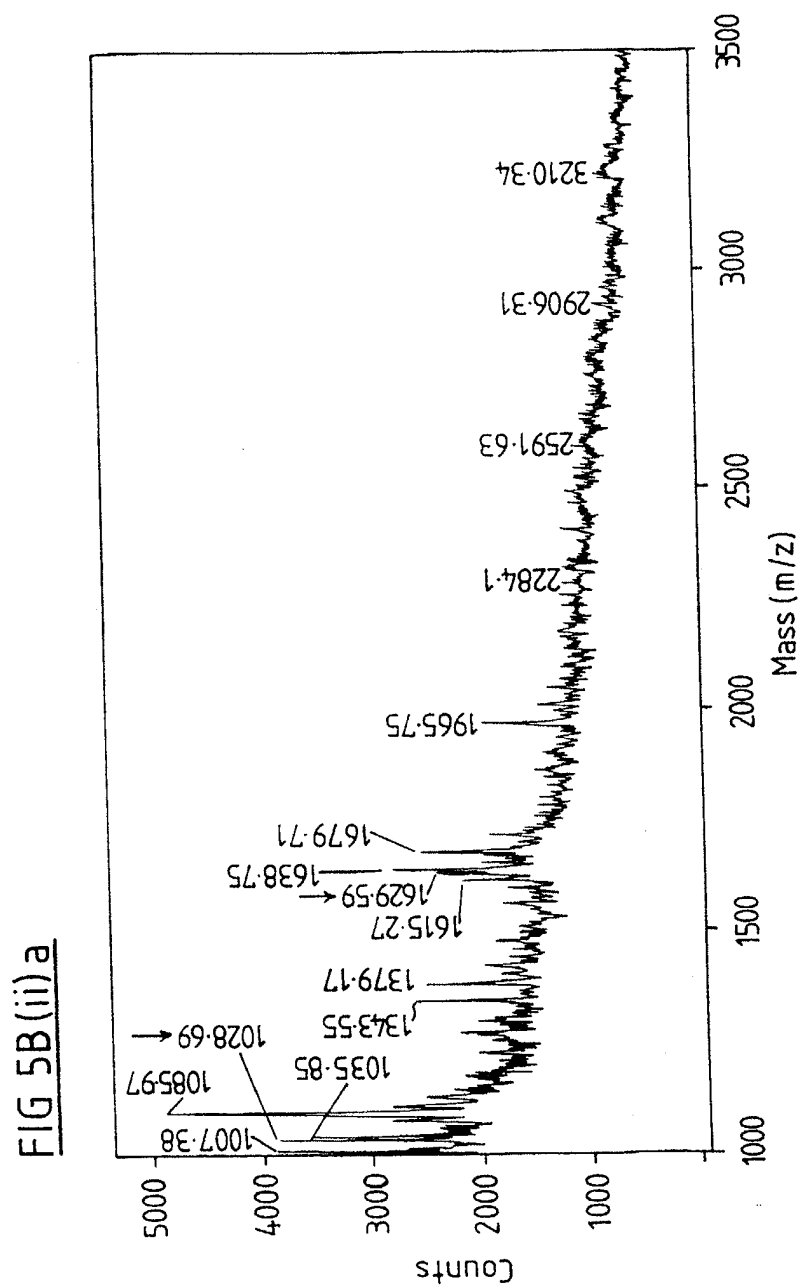
6/13



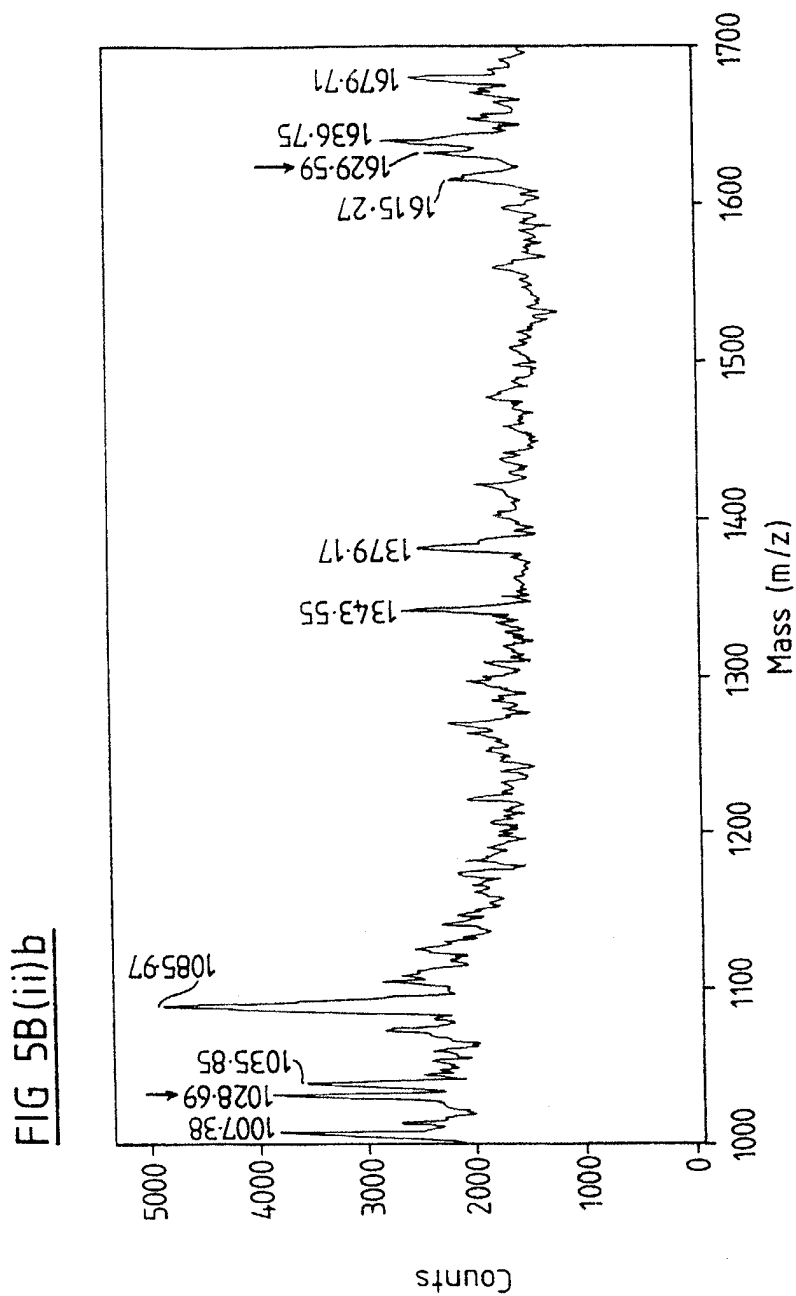
7/13



8/13



9/13



10/13

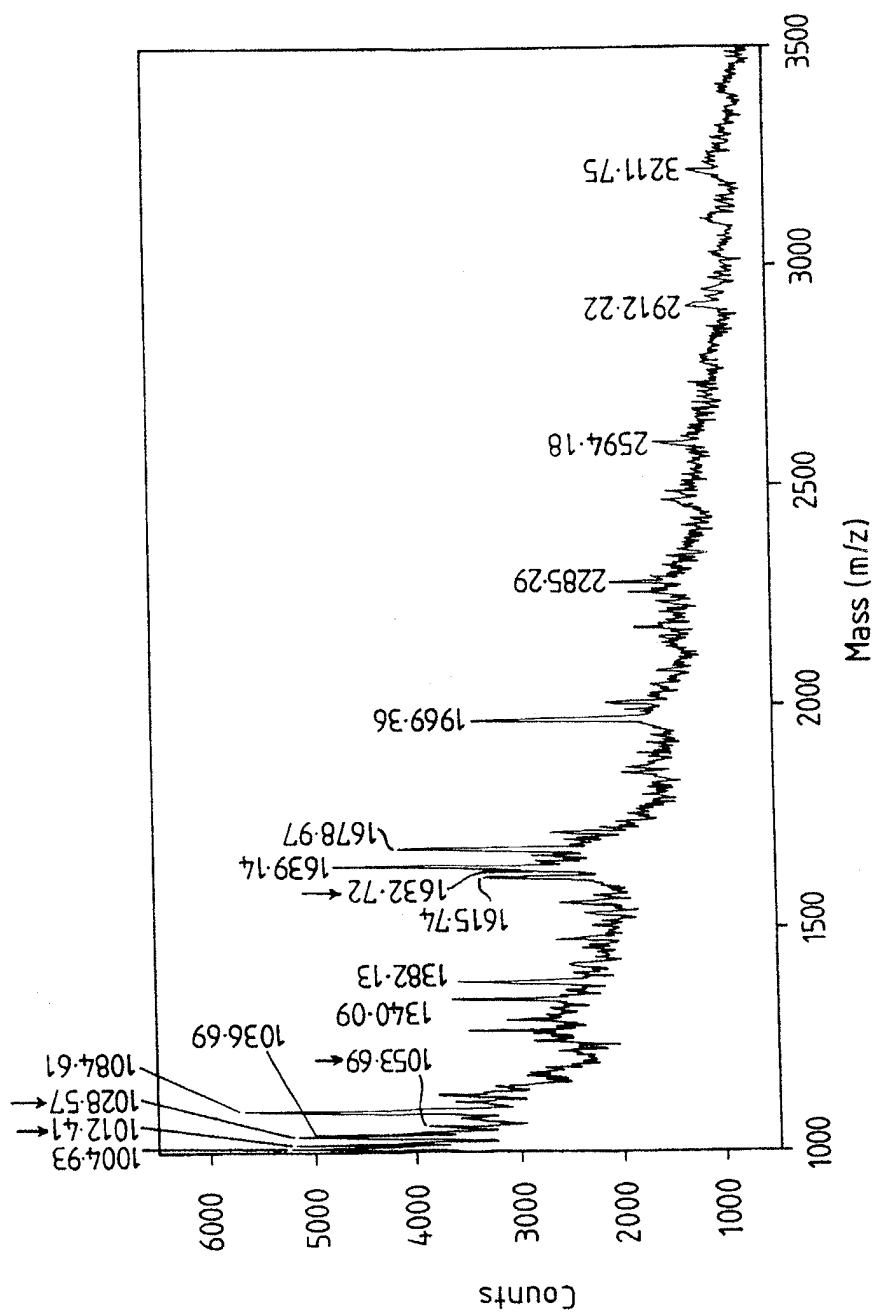


FIG 5B(iii)a

11/13

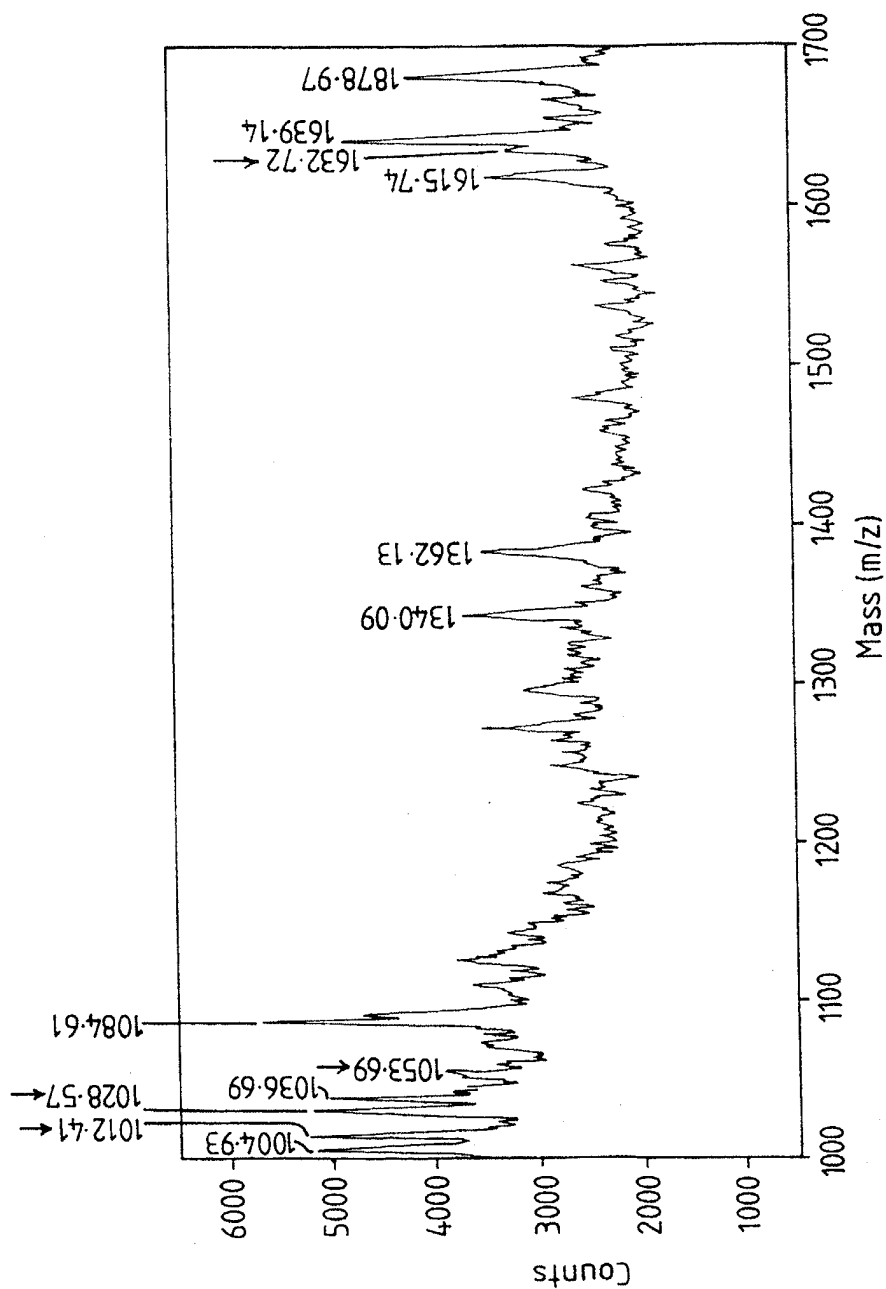
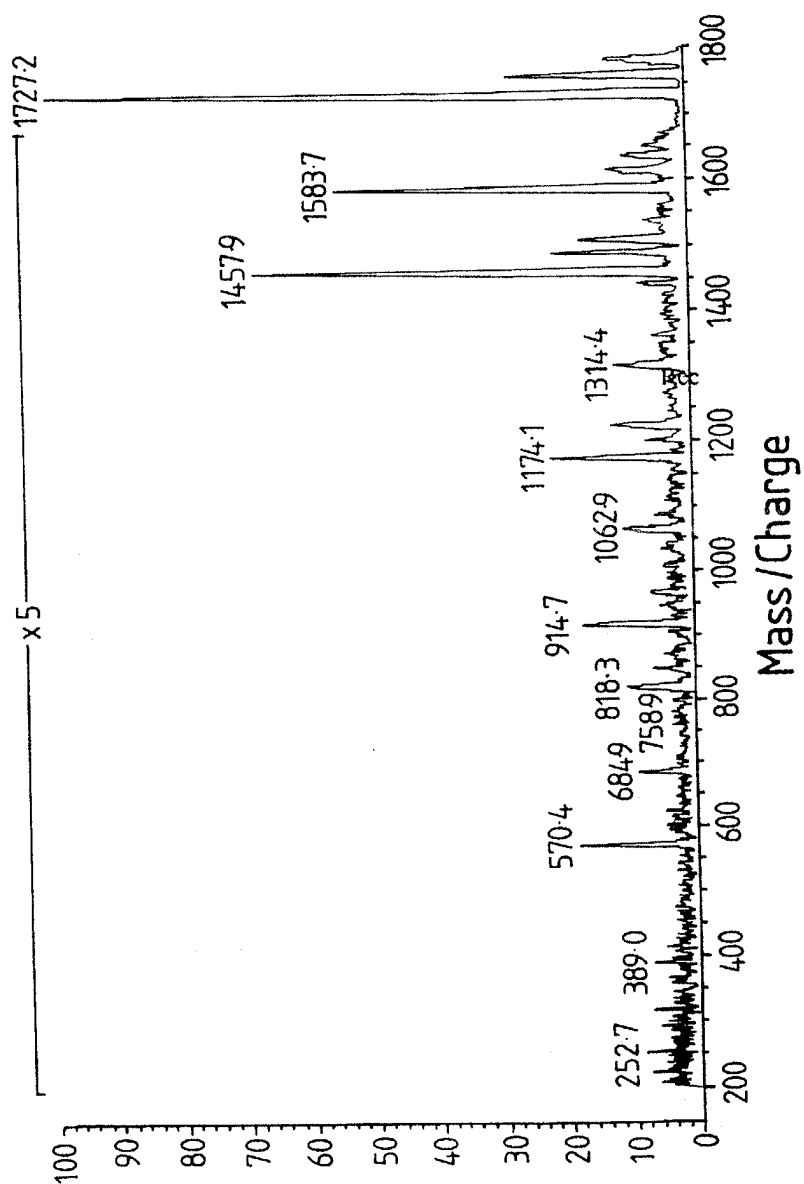
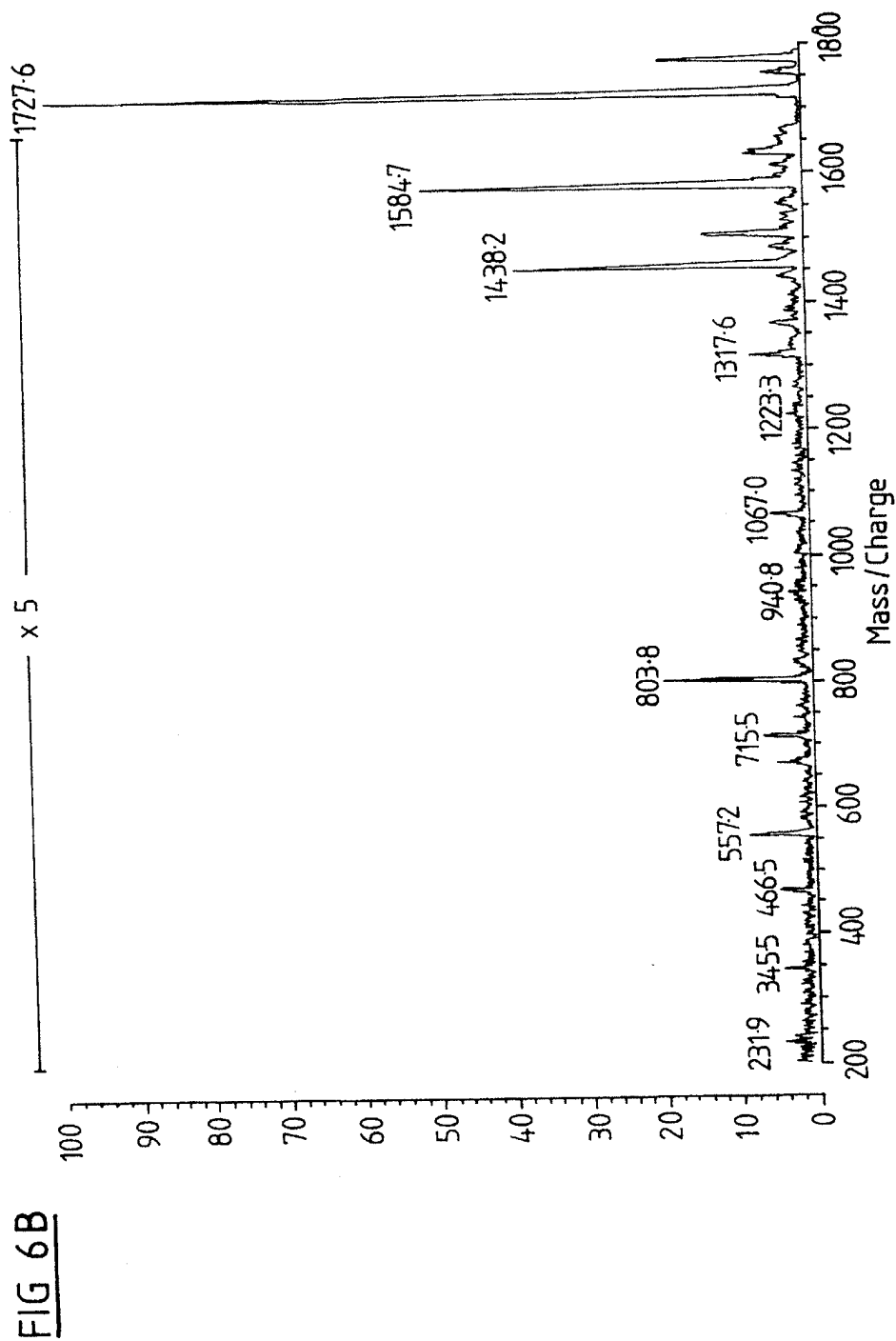


FIG 5B(iii)b

12/13



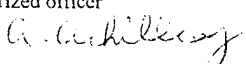
13/13



SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00396

A. CLASSIFICATION OF SUBJECT MATTER		
Int Cl ⁶ : G01N 33/483, C12Q 1/527, 1/68		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC : G01N C12Q and keywords mass spectrom : and oligonucleotide or DNA or nucleic acid		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) DERWENT (WPAT) and keywords as above CHEMICAL ABSTRACTS (CA) and keywords as above		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96/29431 A (SEQUENOM, INC) 26 September 1996 page 5 lines 13-21, page 6 lines 8-13, figs 7B, 9, 21A-C and claims 33 and 48	1-32
X	WO 96/36986 A (PERSEPTIVE BIOSYSTEMS, INC) 21 November 1996 whole document	1-32
X	WO 97/33000 A (GENETRACE SYSTEMS) 12 September 1997 page 10 line 1 to page 13 line 22, examples, 5, 8 and 11	1-32
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C <input checked="" type="checkbox"/> See patent family annex		
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search 7 August 1998		Date of mailing of the international search report 02.09.98
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929		Authorized officer  ANDREW ACHILLEOS Telephone No.: (02) 6283 2280

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00396

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 98/12355, A (GENETRACE SYSTEMS) 26 March 1998 page 4 line 13 to page 6 line 32, page 9 lines 21 to page 10 line 7, examples 5 to 7	1-32
P,X	WO 98/20166 A (SEQUENOM, INC) 14 May 1998 abstract and claims	1-32

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.
PCT/AU 98/00396

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report				Patent Family Member			
WO	96/29431	AU	53651/96	CA	2214359	EP	815261
		US	6505798				
WO	96/36986	EP	827628				
WO	97/33000	AU	20695/97				
WO	98/12355	AU	45916/97				
WO	98/20166	AU	51069/98				
END OF ANNEX							

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C07H 21/02, 21/04, C07K 15/00, C12Q 1/68, C12P 19/34		A1	(11) International Publication Number: WO 94/21663 (43) International Publication Date: 29 September 1994 (29.09.94)
(21) International Application Number: PCT/US94/03246 (22) International Filing Date: 24 March 1994 (24.03.94) (30) Priority Data: 08/036,481 24 March 1993 (24.03.93) US 08/181,629 13 January 1994 (13.01.94) US (71) Applicant: MOLECULAR BIOLOGY RESOURCES, INC. [US/US]; 5520 West Burleigh Street, Milwaukee, WI 53210 (US). (72) Inventors: MEAD, David; 2301 Waltham Road, Madison, WI 53711 (US). SWAMINATHAN, Neela; 476 Presidential Lane, Madison, WI 53711 (US). VAN ETEN, James; 7409 Stevens Ridge Road, Lincoln, NE 68516 (US). SKOWRON, Piotr; 7400 W. Center Street, Wauwatosa, WI 53210 (US). (74) Agent: CLOUGH, David, W.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 223 S. Wacker Drive, Chicago, IL 60606-6402 (US).			(81) Designated States: AU, CA, FI, JP, SE, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: DINUCLEOTIDE RESTRICTION ENDONUCLEASE PREPARATIONS AND METHODS OF USE (57) Abstract <p>The present invention is directed to materials and methods for the quasi-random and complete fragmentation of DNA using restriction endonuclease reagents capable of cutting DNA at a dinucleotide sequence. The invention is also directed to methods for labeling DNA, for shotgun cloning, for sequencing of DNA, for epitope mapping and for anonymous primer cloning, all using fragments of DNA generated by the method of the present invention. In addition, the present invention is directed to DNA sequences encoding a novel restriction endonuclease (designated R. Cvi II) and variants thereof as well as to methods and materials for production of the same by recombinant methods. A bacterial host cell transformed with DNA encoding R. Cvi II is also disclosed as well as methods for expressing R. Cvi II in the bacterial host system and subsequent materials and methods for purifying the enzyme.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MD	Republic of Moldova	TJ	Tajikistan
DE	Germany	MG	Madagascar	TT	Trinidad and Tobago
DK	Denmark	ML	Mali	UA	Ukraine
ES	Spain	MN	Mongolia	US	United States of America
FI	Finland			UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

DINUCLEOTIDE RESTRICTION ENDONUCLEASE PREPARATIONS AND METHODS OF USE.

FIELD OF THE INVENTION

5 The present invention relates generally to isolated purified polynucleotides which encode restriction enzymes and to methods of expressing the restriction enzymes from such polynucleotides. More particularly this invention relates to isolated purified polynucleotides which encode CviJI and related methods for the production of this enzyme.

10 Other aspects of the invention relate to methods for partially or completely digesting DNA at a dinucleotide sequence. More particularly, this aspect of the invention relates to methods of generating quasi-random fragments of DNA, and methods of cloning, labeling, and sequencing DNA, as well as epitope mapping of proteins. The invention also relates to methods for generating
15 sequence-specific oligonucleotides from DNA, without prior knowledge of the nucleic acid sequence of such DNA, and to methods for cloning and labeling DNA after restriction digestion by a two base recognition endonuclease reagent. This invention also relates to methods for cloning, labeling, and detecting nucleic acids using two base restriction endonuclease reagents, such as CviI I, BsuR I,
20 Aci I or CGase I. Further the invention relates to labeling DNA by taking advantage of certain properties of the holo-enzyme of thermostable DNA polymerases.

BACKGROUND OF THE INVENTION

25 Restriction endonucleases are a group of enzymes originally found to be expressed in a wide variety of prokaryotic organisms. More recently they have also been found to be encoded in viral genomes. These enzymes catalyze the selective cleavage of DNA at generally short sequences, often unique to the individual enzyme. This ability to cleave makes restriction endonucleases indispensable tools in recombinant DNA technology. The increased commercial

- 2 -

availability of the isolated enzymes has contributed in large part to the enormous expansion in the field of recombinant DNA technology over the last few years.

These enzymes have been classified into three groups. Because of properties of the type I and type III enzymes, they have not been widely used in molecular biology applications, and will not be discussed further. Type II enzymes are part of a binary system known as a restriction modification system consisting of a restriction endonuclease that cleaves a specific sequence of nucleotides and a separate DNA modifying enzyme that modifies the same recognition sequence and thereby prevents cleavage by the cognate endonuclease. A total of about 2103 restriction enzymes are known, encompassing 179 different type II specificities (Roberts, *et al.*, *Nucl. Acids Res.* 20:2167-2180 (1992)). Although there are more than 1200 type II restriction enzymes, many of them are members of groups which recognize the same sequence. Restriction enzymes that recognize the same sequence are said to be isoschizomers.

The vast majority of type II restriction enzymes recognize specific double-stranded sequences which are four, five, or six nucleotides in length and which display twofold (palindromic) symmetry. A few enzymes recognize longer sequences or degenerate sequences.

The location of cleavage sites within a palindrome differs from enzyme to enzyme. Some enzymes cleave both strands exactly at the axis of symmetry generating fragments of DNA that carry blunt ends, while others cleave each strand at similar sequences on opposite sides of the axis of symmetry, creating fragments of DNA that carry protruding, single-stranded termini.

Restriction endonucleases with shorter recognition sequences cut DNA more frequently than those with longer recognition sequences. For example, assuming a 50% G-C content, a restriction endonuclease with a 4-base recognition sequence will cleave, on average, every 4^4 (256) bases compared to every 4^6 (4096) bases for a restriction endonuclease with a 6-base recognition sequence. Under certain conditions some restriction endonucleases are capable

- 3 -

of cleaving sequences which are similar but not identical to their defined recognition sequence. This altered specificity has been termed "star" (*) activity and is observed only under certain non-standard reaction conditions. The manner in which an enzyme's specificity is altered depends on the particular enzyme and on the conditions employed to induce the star activity. Conditions that contribute to star activity include high glycerol concentration, high ratio of enzyme to DNA, low ionic strength, high pH, the presence of organic solvents, and the substitution of Mg^{++} with other divalent cations. The most common types of star activity involve cutting at a recognition sequence having a single base substitution, cutting at sites having truncation of the outer bases of the recognition sequence, and single-strand nicking. The following restriction endonucleases show star activity: Ase I, BamH I, BssH II, BsuR I, CviJ I, EcoR I, EcoR V, Hind III, Hinf I, Kpn I, Pst I, Pvu II, Sal I, Sca I, Taq I, and Xmn I. Star activity is generally viewed as undesirable, and of little intrinsic value.

Of the 179 unique type II restriction endonucleases, 31 have a 4-base recognition sequence, 11 have a 5-base recognition sequence, 127 have a 6-base recognition sequence, and 10 which have recognition sequences of greater than 6 bases. In two cases, a restriction endonuclease has a recognition sequence of less than 4 bases.

The restriction enzyme CviJ I has a three base recognition sequence or a two-base recognition sequence, depending on the reaction conditions. Under normal reaction conditions CviJ I recognizes the sequence PuGCPy (wherein Pu=purine and Py=pyrimidine) and cleaves between the G and C to leave blunt ends (Xia et al., 1987. *Nucleic Acids Res.* 15:6075-6090). Under "relaxed" or "star" conditions (in the presence of 1 mM ATP and 20 mM DTT) the specificity of CviJ I may be altered to cleave DNA more frequently. This activity is referred to as CviJ I*, for star or altered specificity. However, CviJ I* activity is not observed under conditions which favor star activity of other restriction endonucleases.

- 4 -

The restriction enzyme BsuR I normally recognizes the sequence GGCC and cleaves between the G and C to leave blunt ends. (Heininger, *et al.*, *Gene* 1:291-303 (1977)). Under relaxed conditions (high pH, low ionic strength, and high glycerol concentration) the specificity of Bsu RI may be altered to cleave
5 DNA more frequently. An isoschizomer of this enzyme, Hae III, does not display this star activity.

In bacteria, the restriction endonuclease provides a mechanism of defense against foreign DNA molecules (e.g., bacteriophage DNA) by virtue of its ability to distinguish and cleave only exogenous DNA, leaving endogenous
10 bacterial DNA unaffected. Viral endonucleases possess the same discerning capabilities, but rather than providing a means for defense, this activity has presumably evolved to cripple the host's ability to replicate its own DNA and allows the virus to assume control of the host's replication machinery.

Bacteria and viruses which express restriction endonucleases
15 necessarily possess the inherent ability to protect their own genome from cleavage by their endogenous endonuclease. The primary mechanism by which this is accomplished is by modifying the organisms own DNA by, for example methylating a base in the recognition sequence which prevents binding and cleavage by the endonuclease. Therefore, to insure viability, the genome of an
20 organism which expresses a restriction endonuclease is almost always heavily modified, usually by methylation of cytosine or adenosine bases. The methylase enzyme which modifies the genome (itself a useful tool in molecular biology) acts in tandem with the endonuclease, either as part of an enzyme complex (restriction/modification complex) or as two distinct entities. Therefore,
25 recognizing that an organism expresses an enzyme with endonuclease activity strongly suggests the expression of an associated modifying methylase enzyme (and vice versa) and this association has led to isolation and cloning of a number of commercially available restriction/modification enzymes for use in the laboratory as discussed below.

- 5 -

One of the limitations in the use of restriction endonucleases exists when cleavage of a given sequence is required and no known endonuclease exists which is specific for that particular sequence. Therefore, the continued identification and isolation of unique restriction endonucleases and altered reaction conditions will allow for even more sophisticated manipulation of DNA *in vitro*.

A number of publications and patents describe the cloning of DNAs encoding restriction endonucleases. Included among these publications is Kiss, A., *et al.*, *Nucleic Acid Research* 13:6403-6421 (1985), which describes the cloned nucleotide sequence of the *Bsu*RI restriction-modification system isolated from *Bacillus subtilis*. This system is specific for the sequence 5'-GGCC-3' and is defined by two gene products which are transcribed by different promoters. The methylase component of the system shows homology to the methylase from the *Bsp*RI and *SPR* restriction-modification systems.

Nwanko, D.O. and Wilson, G.G. *Gene* 64:1-8 (1988), describe the cloning and expression of the *Msp*I restriction and modification genes isolated from *Moraxella* sp. This system recognizes the sequence 5'-CCGG-3' and both enzymes are functional in *E. coli*. Evidence indicates that these genes are transcribed in opposite directions, thus are probably under the control of different promoters.

Ashok, K.D., *et al.*, *Nucleic Acids Research* 20:1579-1585 (1992), describe the purification and characterization of cloned *Msp*I methyltransferase, over-expressed in *E. coli*. At low concentrations the enzyme exists as a monomer, but at higher concentrations it exists mainly as a dimer. Polyclonal antibodies to the enzyme cross-react with methyltransferase genes of other modification systems.

Brooks, J.E., *et al.* *Nucleic Acids Research* 19:841-850 (1991), characterizes the cloned *Bam*HI restriction modification system from *Bacillus subtilis*. The two genes are divergently oriented and separated by an open reading frame which may serve as a transcriptional regulator in the native bacteria.

- 6 -

Slatko, B.E., *et al.* *Nucleic Acids Research* 15:9781-9796 (1987), describe the cloning, sequencing and expression of the *TaqI* restriction-modification system. These genes have the same transcriptional orientation, with the methylase gene 5' to the endonuclease gene. *E. coli* clones which carry only
5 the endonuclease gene are viable even in the absence of the methylase gene. This is an unusual case possibly explained by the 65°C optimal temperature for *TaqI* restriction and the 37°C optimal temperature for *E. coli* growth.

Howard, K.A., *et al.*, *Nucleic Acids Research* 14:7939-7951 (1986), describe the cloning of the *DdeI* restriction modification system from
10 *Desulfovibrio desulfuricans* by a two step method wherein the methylase gene is first cloned and transformed into *E. coli*, followed by the cloning of the endonuclease gene and transformation of this second gene into the methylase-expressing bacteria. In order to maintain cell viability, high levels of methylase expression are required before the endonuclease gene can be introduced into the
15 bacteria.

Ito, H., *et al.*, *Nucleic Acids Research* 18:3903-3911 (1990), describe the cloning, nucleotide sequence and expression of the *HincII* restriction-modification system. The DNA was isolated from *H. influenzae* Rc, with the two genes positioned in the same transcriptional orientation.

Shields, S.L., *et al.*, *Virology* 76:16-24 (1990), describe the
20 cloning and sequencing of the cytosine methyltransferase gene M.CviII from the *Chlorella* virus IL-3A. The methylase recognizes the sequence (G/A)GC(T/C/G) and shows amino acid sequence homology with 5-methylcytosine methylases isolated from bacteria. DNA encoding the methylase was obtained from the viral
25 genome which was propagated in the green alga host *Chlorella*.

Xia, Y., *et al.*, *Nucleic Acids Research* 15:6075-6090 (1987), discovered that IL-3A virus infection of *Chlorella*-like green alga induces the expression of the DNA restriction endonuclease *CviII* which has novel sequence specificity. This endonuclease recognizes the sequence PuGCPy (wherein Pu =

- 7 -

purine and Py = pyrimidine) but does not cut the sequence PuG^mCPy, where ^mC is 5-methylcytosine.

U.S. Patent 5,137,823, issued August 11, 1992, to Brooks, J.E., describes a two step method for cloning the *Bam*HI restriction modification system wherein the methylase is cloned first and then introduced into a bacterial host. The endonuclease is then cloned and introduced into the methylase expressing bacteria. This two step procedure provides the host DNA protection from cleavage of the subsequently introduced endonuclease.

U.S. Patent 5,200,333, ('333) issued April 6, 1993, to Wilson, G.G., describes a method for cloning restriction and modification genes. Specifically this reference describes the cloning of the *Taq*I and *Hae*II systems from *Thermus aquaticus* and *Haemophilus aegypticus*, respectively. In this method, bacterial DNA was initially purified and digested, and the fragments were then cloned into a vector to produce a bacterial DNA library. The library was then transformed into *E. coli* and the cells were plated. Colonies were then scraped from the plate to form a primary cell library. Plasmid DNA from this cell library was purified and digested with the endonuclease of the two gene system. Bacteria which expressed the methylase gene had modified plasmid DNA which was protected from endonuclease activity, while plasmids from bacteria which lacked the intact methylase gene were digested. The resulting, undigested plasmid DNA was then transformed into another bacterial strain and the bacteria were plated. Surviving colonies were again harvested to give a secondary cell library and the entire procedure repeated. Plasmids which code for the complete restriction-modification system presumably survived each round of purification and were enriched. Bacteria which survive several rounds of enrichment were subsequently assayed for both methylase and endonuclease activity.

U.S. Patent 5,196,331, ('331) issued March 23, 1993, to Wilson, G.G. and Nwanko, D., describes a method for cloning the *Msp*I restriction and modification genes. This patent describes a method identical to that of U.S.

- 8 -

Patent 5,200,333 ('333). '331 is a continuation-in-part of, and '333 is a continuation of U.S.S.N. 707,079 (now abandoned).

As mentioned above, *Chlorella* virus IL-3A encodes a unique restriction endonuclease called CviJI (Xia *et al.* *Nucleic Acids Res.* 15:6075-6090 (1987)). IL-3A is a large, polyhedral, plaque-forming phycodnavirus (Francki, R.I.B., *et al.* *Arch. Virol.* suppl.2. Springer-Verlag, Vienna (1991)) that replicates in unicellular, eukaryotic green algae, *Chlorella* strain NC64A (Schuster, A.M., *et al.* *Virology* 150:170-177 (1986)). The double-stranded DNA genome of IL-3A is approximately 330 kbp (Rohozinski *et al.*, *Virology* 168:363-369 (1989)) and contains 9.7% methylated cytidine (Van Etten, J.L. *et al.*, *Nucleic Acids Res.* 13:3471-3478 (1985)). The cognate methyltransferase of CviJI, M.CviJI, methylates (A/G)GC(T/C/G) sequences and, has been cloned and sequenced (Shields, S.L. *et al.*, *Virology* 176:16-24 (1990)).

The use of a two/three base recognition endonuclease, such as CviJI, to improve numerous conventional molecular biology applications as well as permitting novel applications has been described in co-pending U.S. Patent Application Ser.No. 08/036,481, filed on March 24, 1993. The application discloses methods for generating sequence-specific oligonucleotides from DNA without prior knowledge of the nucleic acid sequence of such DNA, and to methods for cloning and labeling DNA after restriction digestion by a two base recognition endonuclease. The application also teaches methods for generating quasi-random fragments of DNA, methods for cloning, labeling, and sequencing DNA, as well as epitope mapping of proteins. The ability to generate numerous oligonucleotides with perfect sequence specificity or quasi-random distributions of DNA fragments such as is possible with CviJI* has important implications for a number of conventional and novel molecular biology procedures.

Infection of *Chlorella* species NC64A with the IL-3A virus produces sufficient CviJI restriction endonuclease (CviJI) for research purposes. However, production of commercially useful amounts of CviJI is limited with this

- 9 -

system due to the slow growth of *Chlorella* algae, the large number of contaminating nucleases associated with the virus, and the small yield of enzyme obtained after purification. In addition, biochemical and biophysical characterization of the enzyme, such as molecular weight determination, are difficult from the native source. Because of these limitations it would be useful to clone the gene for CviJI in order to provide an adequate large scale source of enzyme for use as a molecular biological reagent.

SUMMARY OF THE INVENTION

In one of its aspects, the present invention provides purified and isolated polynucleotides (e.g., DNA sequences and RNA transcripts thereof) encoding a unique restriction endonuclease, CviJI, as well as polypeptides and variants thereof which display activities characteristic of CviJI. Activities of CviJI include the recognition of specific DNA sequences, binding to these sequences and cleaving the bound DNA into fragments. Preferred DNA sequences of the invention include viral genomic sequences as well as wholly or partially chemically synthesized DNA sequences. Replicas (i.e., copies of the isolated DNA sequences made *in vivo* or *in vitro*) of DNA sequences of the invention are also contemplated. A preferred DNA sequence is set forth in SEQ ID NO: 2 herein and is contained as an insert in the plasmid pCJH1.4. In another of its aspects, the invention provides purified isolated DNA encoding a CviJI polypeptide by means of degenerate codons.

Also provided are autonomously replicating recombinant constructions such as plasmid DNA vectors incorporating CviJI sequences and especially vectors wherein DNA encoding CviJI or a CviJI variant is operatively linked to an endogenous or exogenous expression control DNA sequence.

According to another aspect of the invention, host cells such as prokaryotic and eukaryotic cells, are stably transformed with DNA sequences of the invention in a manner allowing the desired polypeptides to be expressed

- 10 -

therein. Host cells expressing *Cvi*II and *Cvi*II variant products are useful in methods for the large scale production of *Cvi*II and *Cvi*II variants wherein the cells are grown in a suitable culture medium and the desired polypeptide products are isolated from the host cells or from the medium in which the cells are grown.

5 A preferred host cell is *E. coli*. Still another aspect of the invention is a recombinant *Cvi*II polypeptide.

The present invention is also directed to a method for the digestion of DNA with a restriction endonuclease reagent under conditions wherein said DNA is cleaved at a dinucleotide sequence selected from the group consisting of
10 PyGCPy, PuGCPy, PuGCPu, and wherein Pu = purine and Py = pyrimidine.

The present invention is also directed to a method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is digested at 11 of 16 possible dinucleotide sequences and wherein said dinucleotide
15 sequences are selected from the group consisting of PuCGPu, PuCGPy, and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

The present invention is directed to shotgun cloning of DNA, epitope mapping, and for labeling DNA using the digestion methods of the present invention. The present invention provides methods for quasi-random fragmenting
20 of DNA using the digestion methods of the present invention under conditions wherein the DNA is only partially cleaved and the site preference of the restriction endonuclease reagent is greatly reduced. By quasi-random is meant an overlapping population of DNA fragments produced by digesting DNA using the methods of the present inventions without apparent site-preference and which
25 appears as a smear upon electrophoresis in a 1-2 wt. % agarose gel. The present invention is also directed to the shotgun cloning and sequencing of quasi-random fragments of DNA produced by the methods of the present invention. Quasi-random fragments in the shotgun cloning method of the present invention are produced by partial digestion of DNA with a restriction endonuclease reagent

- 11 -

according to the methods of the present invention. More particularly, quasi-random fragments of DNA useful in the cloning method of the present invention are produced by the partial digestion of the DNA to be cloned with CviI I, BsuR I or with a restriction endonuclease reagent termed CGase I comprising Taq I and Hpa II. Quasi-random fragments having a length of between about 100 and about 10,000 nucleotides are preferred. More preferred are quasi-random fragments of about 500 to about 10,000 nucleotides in length. The present invention is also directed to the generation of quasi-random fragmentation of DNA using the method of the present invention for the purposes of epitope mapping and gene cloning. These quasi-random fragments are expressed either *in vitro* or *in vivo* and the smallest fragment containing the desired function is identified by screening assays well known in the art.

The present invention is also directed to the production of anonymous primers from any DNA without prior knowledge of the nucleotide sequence. The present invention provides methods for anonymous primer cloning and sequencing after complete digestion of DNA utilizing CviI I, BsuR I or CGase I using the methods of the present invention.

Additionally, the present invention is directed to methods of labeling and detecting DNA comprising the complete digestion of DNA using the methods of the present invention, followed by a heat denaturation step, to yield sequence specific oligonucleotides. In particular, an aspect of the present invention involves labeling DNA with sequence specific oligonucleotides of about 20 to about 200 bases in length (with an average size of between 20-60 bases) generated by CviI I, BsuR I or CGase I digestion of the template DNA.

More particularly, the invention is directed to restriction generated oligonucleotide labeling (RGOL) of DNA which comprises the digestion of an aliquot of template DNA with CviI I followed by a simple heat denaturation step, thereby generating numerous sequence specific oligonucleotides, which can then be utilized for labeling nucleic acids by a number of methods, including primer

- 12 -

extension type reactions with a DNA polymerase and various labels, isotopic or non-isotopic (RGOL-PEL); 5' end labeling with polynucleotide kinase; 3' end labeling using terminal transferase and various labels, isotopic or non-isotopic. Labeling at the 3' end, also referred to as tailing, adds numerous labels per oligonucleotide (1-200), depending on the labeling conditions. The addition of 10-500 oligonucleotides generated per template, results in a significant signal amplification not obtainable by conventional methods.

The invention is also directed to thermal cycle labeling (TCL) which comprises the simultaneous labeling and amplification of probes utilizing CviI or CGase I restriction generated oligonucleotides as the starting material. In this method, natural DNA of unknown sequence is digested with CviI to generate numerous double-stranded fragments which are then heat denatured to yield oligonucleotides. These oligonucleotides are combined with the intact template and subjected to repeated cycles of denaturation, annealing, and extension in the presence of a thermostable DNA polymerase or functional fragment thereof which maintains polymerase activity, deoxynucleotide triphosphates and the appropriate buffer. Alpha ^{32}P -dATP (or any of the other three deoxynucleotide triphosphates), biotin-dUTP, fluorescein-dUTP, or digoxigenin-dUTP is incorporated during the extension step for subsequent detection purposes. Thermal cycle labeling efficiently labels DNA while simultaneously amplifying large amounts of the labeled probe. In addition, TCL probes exhibit a 10 fold improvement in detection sensitivity compared to conventional probes.

The present invention is also directed to TCL in which the thermostable DNA polymerase supplies endogenous primers for enzymatic extension. This method is referred to as Universal Thermal Cycle Labeling (UTCL). In this method natural DNA of unknown sequence is combined intact with the holo-enzyme of a thermostable DNA polymerase, deoxyribonucleotide triphosphates, and the appropriate buffer. The holo-enzyme and its associated

- 13 -

endogenous primers are then combined with intact template and subjected to repeated cycles of denaturation annealing and extension. Alpha ^{32}P -dATP, ^{32}P -dTTP, ^{32}P -dGTP, ^{32}P -dCTP, biotin-dUTP, fluorescein-dUTP, or digoxigenin-dUTP is also included in the extension step for subsequent detection purposes.

5 Isotopic labels useful in the practice of the present invention include but are not limited to ^{32}P , ^{33}P , ^{35}S , ^{14}C and ^3H . Non-isotopic labels useful in the present invention include but are not limited to fluorescein biotin, dinitrophenol and digoxigenin.

10 The present invention is also directed to an improved method for purifying CviI I from the algae *Chlorella* infected with the virus IL-3A.

In addition the present invention is directed to restriction endonuclease reagents which, under conditions which relax the sequence specificity of one or more restriction endonucleases, cleave DNA at the dinucleotide sequences AT or TA.

15 The present invention is also directed to a restriction endonuclease reagent comprising in combination, Taq I and Hpa II, which is capable of digesting DNA at 11 of 16 possible dinucleotide sequences, said sequences selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

20 The following examples are intended to be illustrative of the several aspects of the present invention and are not intended in any way to limit the scope of any aspect of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 is a map of the plasmid p710 which contains DNA sequences encoding for the IL-3A viral methyltransferase M.CviII;

Figure 2 is the nucleotide sequence of 5497 bp of cloned IL-3A viral DNA;

- 14 -

Figure 3 is a restriction map of the cloned IL-3A viral DNA, including the identified open reading frames;

Figure 4 is the DNA sequence of the *Cvi*II gene with its flanking regions. The predicted amino acid sequence is provided below the nucleotide sequences;

Figure 5A depicts the theoretical frequency and distribution of *Cvi*II* restriction generated oligomers of individual lengths; Figure 5B shows the actual frequency and distribution of *Cvi*II* restriction generated oligomers of various lengths;

Figure 6 is a flow chart depicting anonymous primer cloning;

Figure 7 is a photographic reproduction of a gel depicting *Cvi*II restriction digests of pUC19;

Figure 8 is a photographic reproduction of a gel depicting comparisons of sonicated versus *Cvi*II* partially digested DNAs;

Figure 9A is a photographic reproduction of an agarose gel electrophoresis analysis of size-fractionated DNA by microcolumn chromatography compared to fractionation by agarose gel electroelution;

Figure 9B-E illustrates additional trials of the same procedures used in Figure 9A;

Figure 10A illustrates the size distribution of DNA fragments produced by partial digestion of DNA by *Cvi*II and fractionated by microcolumn chromatography;

Figure 10B-C illustrates the size distribution of DNA fragments produced by partial digestion of DNA by *Cvi*II and fractionated by agarose gel electrophoresis;

Figure 11 is a schematic depiction of the distribution of *Cvi*II sites in pUC19; and

Figure 12 is a graph of the rate of sequence accumulation by *Cvi*II** shotgun cloning and sequencing.

- 15 -

DETAILED DESCRIPTION

The gene for the restriction endonuclease R.CviJI was cloned into *E. coli* so as to provide an adequate source of R.CviJI for use as a molecular biological reagent. Biologically active CviJI has been purified from *E.coli* to
5 apparent homogeneity. The molecular weight of *E.coli* derived R.CviJI is 32.5 kD by SDS gel electrophoresis. N-terminal amino acid sequence analysis of this protein and comparison to the nucleotide sequence of the gene revealed that the translation of this enzyme is probably initiated with a GTG start codon, instead of the usual ATG initiation codon. The structural gene is 834 nucleotides in
10 length coding for a protein of 278 amino acids (31.6 kD). A second peak of R.CviJI activity which elutes separately from the 32.5 kD form can be seen in the initial stages of enzyme purification. Trace amounts of a larger molecular weight form have not been observed to date. However, the R.CviJI gene does possess an in-frame upstream ATG codon which if translated would yield a predicted 41.4
15 kD protein. The structural gene for this potentially larger product is 1074 nucleotides in length coding for a putative protein of 358 amino acids.

The present invention is also directed to a method for the fragmentation and cloning of DNA using the restriction endonuclease CviJI under conditions which allow the enzyme to cleave DNA at the dinucleotide sequence
20 GC. In addition, the present invention is also directed to the cloning of quasi-random fragments of DNA digested using the fragmentation method of the present invention.

As an alternative to the methods for constructing random clone libraries described above, methods were devised for the construction of such
25 libraries which require fewer steps and reagents, which require smaller amounts of DNA, which have relatively high cloning efficiencies and which takes less time to complete. These methods relate to the recognition that a partial digest with a two or three base recognition endonuclease cleaves DNA frequently enough to be functionally random with respect to the rate at which sequence data may be

- 16 -

accumulated from a shotgun clone bank. The restriction enzyme CviI I normally recognizes the sequence PuGCPy and cleaves between the G and C to leave blunt ends (Xia *et al.*, *Nucl. Acids Res.* 15:6075-6090 (1987)). Under "relaxed" conditions (in the presence of 1 mM ATP and 20 MM DTT) the specificity of CviI I can be altered to cleave DNA more frequently and perhaps as frequently as at every GC. This activity is referred to as CviI I*. Because of the high frequency of the dinucleotide GC in all DNA (16 bp average fragment size for random DNA), quasi-random libraries may be constructed by partial digestion of DNA with CviI I*. A DNA degradation method with low levels of sequence specificity produces a smear of the target DNA when analyzed by agarose gel electrophoresis. Digestion of the plasmid pUC19 under partial CviI I* conditions does not result in a non-discrete smear; rather, a number of discrete bands are found superimposed upon a light background of smearing, suggesting that CviI I* has some site preference. Atypical reaction conditions according to the present invention eliminate this apparent site preference of CviI I* to produce an activity (termed CviI I**) in combination with a rapid gel filtration size exclusion step, streamlines a number of aspects involved in shotgun cloning.

One aspect of the present invention involves the use of the two/three base recognition endonuclease CviI I, in conjunction with a simple spin-column method to produce libraries equivalent in final form to those generated by the combination of sonication and agarose gel electroelution. However, the method of the present invention requires fewer steps, a shorter time period, and significantly less substrate (nanogram amounts) when compared to conventional procedures. Both small and large sequencing projects using the methods described herein are within the scope of the present invention.

Current sequencing paradigms require the generation of a new template for each 350-500 nucleotides sequenced. On this basis, sequencing both strands of the human genome would require at least 12 million templates 500 nucleotides long, assuming no overlap between templates.

- 17 -

A random approach, such as shotgun sequencing, would require 30 to 50 million templates, assuming the entire genome were randomly subcloned. As many as 250,000 libraries may be needed to generate the requisite templates from a subcloned and ordered array of this genome, depending on the type of vector utilized, and the degree of overlap between such clones. The ability to generate shotgun libraries in a semi-automated, microtiter plate format would greatly simplify such large scale projects.

The development of methods for cloning large DNA molecules in yeast artificial chromosomes (Burke *et al.*, *Science* 236:806-812 (1987), or in bacteriophage P1-derived vectors (Sternberg, *Proc. Natl. Acad. Sci. USA* 87:103-107 (1990)), simplifies the subdivision and analysis of very large genomes. However, the large size of the resulting subclones (100 - 1000 kbp) presents additional challenges for subsequent sequencing efforts. A report of the sequencing of a 134 kbp genome by random shotgun cloning directly into a bacteriophage M13 vector indicates that numerous intermediate stages of subcloning, mapping, and overlapping such clones may be eliminated (Davison, *J. DNA Seq. and Mapping* 1:389-394 (1992). An order of magnitude reduction in the amount of DNA required for shotgun cloning would substantially simplify efforts to directly sequence 100,000 bp sized molecules and beyond.

The ability to generate an overlapping population of randomly fragmented DNA molecules is considered essential for minimizing the closure of nucleotide sequence gaps by the shotgun cloning method. The use of a very frequent-cutting restriction enzyme, such as CviI I, is an approach which has not been utilized. Reaction conditions according to the present invention result in the quasirandom restriction of pUC19 and lambda DNA, as judged by the degree of smearing observed.

The randomness of this CviI I^{**} reaction was quantified by sequence analysis of 76 such partially-fragmented pUC19 subclones. The analysis is showed that CviI I^{**} partial digestion (limiting enzyme and time) restricts DNA

- 18 -

at PyGCPy, PuGCPu, and PuGCPy (but not PyGCPu), and is thus a hybrid reaction which combines the three base recognition specificity of CviI I with the "two" base recognition specificity of CviI I*. Interestingly, most of the "relaxed" cleavage observed under CviI I** conditions occurred in those portions of the sequence which were deficient in "normal" restriction sites. CviI I** treatment produces a relatively uniform size distribution of DNA fragments, permitting sequence information to be accumulated in a statistically random fashion.

Shotgun cloning with CviI I** digested DNA is efficient partly because the resulting fragments are blunt ended. Other methods currently used to randomly-fragment DNA, including sonication, DNase I treatment, and low pressure shearing, leave ragged ends which must be converted to blunt ends for efficient vector ligation. Other than a heat denaturation step to inactivate the endonuclease, no additional treatments are required for cloning CviI I** restricted DNA. In addition, the preligation step required to equalize representation of the ends of a DNA molecule prior to sonication or DNase I treatment is not necessary with CviI I** fragmentation. CviI I* cleaves its cognate recognition site very close to the ends of a linear molecule, as judged by the very small fragments resulting from complete digestion of pUC19 as depicted in Figure 2, lane 1.

The overall efficiency of shotgun cloning depends not only on the fragmentation process, but also upon the size fractionation procedure used to remove small DNA fragments. The efficiency of cloning agarose gel fractionated DNA was found to be unexpectedly variable. Numerous experiments produced an erratic distribution of sized material and the resulting cloned inserts were uniformly small (70% < 500 bp in one trial, 100% < 500 bp in another). The method of the present invention includes a simple and rapid micro-column fractionation method, which has resulted in three to thirteen times more transformants than agarose gel fractionation. More importantly, the size distribution of the cloned inserts from column-fractionated DNA was skewed

- 19 -

toward larger fragments (88% > 500 bp). Micro-column fractionation also eliminates the chemical extraction steps required for agarose fractionated DNA. After the target DNA has been column-fractionated, no further treatments are required for cloning. Combining CviI^{**} partial restriction with micro-column
5 fractionation permits the construction of useful libraries from as little as 200 ng of substrate, an order of magnitude less starting material than recommended for sonication/end-repair and agarose gel fractionation procedures.

The CviI^{**} reaction represents a unique alternative for controlling the partial digestion of DNA, a technique which is fundamental to the construction
10 of genomic libraries (Maniatis *et al.* *Cell* 15:687-701 (1978), and restriction site mapping of recombinant clones (Smith, *et al.* *Nucl. Acids Res.* 3:2387-2398 (1976). Partial DNA digests are notably variable and are strongly dependent on the concentration and purity of the DNA, the amount of enzyme used, the incubation time, and the batch of enzyme. Partial digestions may also be variable
15 with respect to the rate at which a particular recognition sequence is cleaved throughout the substrate. Optimal reaction conditions, such as those which render such partial digests independent of one or more of these variables, allows more precise control of the end product. Several controlling schemes may be employed, including: the addition of a constant amount of carrier DNA (Kohara
20 *et al.*, *Cell* 50:495-508 (1987)), the use of limiting amounts of Mg²⁺ (Albertson *et al.* *Nucl. Acids Res.* 17:808 (1989)), ultraviolet irradiation (Whitaker, *et al.* *Gene* 41:129-134), and the combination of a restriction enzyme and a sequence complementary DNA methylase (Hoheisel *et al.*, *Nucl. Acids Res.* 17:9571-9582 (1989)). Utilizing three different batches of CviI I, and three different DNA
25 templates from five separate preparations, a uniform CviI I^{**} partial digestion pattern was obtained that was primarily time-dependent when a constant ratio of 0.3 units of enzyme per μ g of DNA was used.

The rate at which a particular restriction site is cleaved at different locations in a substrate is variable for many endonucleases (Brooks, *et al.*,

- 20 -

Methods in Enzymol, 152:113-129 (1987)). Reaction conditions for CviI I may be optimized to substantially reduce the site preferences of this enzyme during partial digestion (see Figure 2, lanes 3 and 4). Normally, "star" reaction conditions result in cleavage at new sites. The use of star reaction conditions according to the present invention (dimethyl sulfoxide [DMSO] and lowered ionic strength) to affect the partial digestion activity of CviI I** does not result in an altered restriction site cleavage as assayed by sequencing the products of 76 digestion reactions. Instead, the relative rate of cleavage of individual sites appears to be more uniform under these conditions. A 3-5 fold increase in the rate of normal CviI I restriction with the standard buffer and DMSO further substantiates this approach. All of these results indicate that, under the appropriate reaction conditions, CviI I is useful for a number of other applications, such as high resolution restriction mapping and fingerprinting, diagnostic restriction of small PCR fragments, and construction of genomic DNA libraries.

Another aspect of the present invention involves quasi-random fragmentation of DNA using the method of the present invention for epitope mapping and cloning intact genes. The same method as described above for shotgun cloning is utilized, except that an expression vector is used to generate functional proteins from the DNA.

Another aspect of the present invention involves fragmenting DNA using the present invention to generate multiple oligonucleotides from any double-stranded DNA template. Restriction-generated oligonucleotides (RGO) are sequence specific oligonucleotides generated from any DNA according to the present invention. CviI I* presumably cleaves the recognition sequence GC between the G and C to leave blunt ends (Xia *et al.*, *Nucl. Acids Res.* 15:6075-6090, (1987)). Because of the high frequency of dinucleotide GC in all DNA (16bp average fragment size for random DNA), a complete CviI I* restriction results in numerous fragments which are about 20-200 bp in size. These

- 21 -

restriction fragments are generated from an aliquot of the template itself and are heat-denatured to yield numerous single-stranded oligonucleotides which are of variable length but which are specific for the cognate template. Complete CviI I* restriction of the small plasmid pUC19 (2689 bp) theoretically yields 314 oligonucleotides after a heat-denaturation step. The ability to generate numerous oligonucleotides with perfect sequence specificity is an unusual result of the use of this class of enzyme according to the present invention. Such oligonucleotides are uniquely suited for purposes of labeling DNA, as described below.

One application of CviI I* restriction-generated oligonucleotides is to directly label them using conventional methods. There are several important advantages in using CviI I* restriction-generated oligonucleotides. Conventional methods employing synthetic oligonucleotides for detection purposes generally use one oligonucleotide containing one or a few labels. A complete CviI I* digest generates hundreds of oligonucleotides from a given template, depending on the size of the template, and thus makes hundreds of sites available for labeling, regardless of the labeling scheme utilized. These hundreds of sequence specific restriction-generated oligonucleotides have two important advantages over conventional probes used in nucleic acid detection methods. First, the generation of multiple oligonucleotide probes directed at multiple sites in a given target (theoretically, 314 sites in pUC19) provides enhanced detection sensitivities compared to synthetic oligonucleotides which are directed at 1 or a few sites in a target. The numerous labeled restriction-generated oligonucleotides represent a 10-100 fold amplification of the signal for detection compared to the use of a single oligonucleotide. Second, the short length of the restriction-generated oligonucleotides permits more efficient hybridization. This is important for two reasons. First, hybridization times using restriction-generated oligonucleotides is reduced to 1 hr as opposed to an overnight incubation with conventional probes hundreds of nucleotides in length. This is a very important advantage when using oligonucleotide probes in clinical settings. Second, the penetration of probes into

- 22 -

permeabilized cells is a critical issue for *in situ* hybridization procedures. The smaller the probe, the easier the entry into the cell. Thus, the use of multiple oligonucleotide probes generated by the two base cutters greatly improves the sensitivity of *in situ* hybridization, a technique of considerable importance in research and clinical labs. Finally, when using membrane-based hybridization procedures, only small sections of a target nucleic acid are exposed and available for hybridization. Multiple oligonucleotides derived from a cognate template exhibit better detection sensitivities compared to long probes.

Another application of restriction-generated oligonucleotides for labeling is to employ them as primers in a polymerase extension labeling reaction in conjunction with a repetitive thermal cycling regimen of denaturation, annealing, and extension. Thermal Cycle Labeling (TCL) is a method for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of the labeled probe. The TCL system employs the two base recognition endonuclease CviI I* to generate sequence-specific oligonucleotides from the template DNA itself. These oligonucleotides are combined with the intact template and subjected to repeated cycles of denaturation, annealing, and extension by a thermostable DNA polymerase from, for example, *Thermus flavus*. A radioactive- or non-isotopically-labeled deoxynucleotide triphosphate is incorporated during the extension step for subsequent detection purposes. The amplified, labeled probes represent a very heterogeneous mixture of fragments, which appears as a large molecular weight smear when analyzed by agarose gel electrophoresis. Primer-primer amplification, a side product of this reaction (produced by leaving out the intact template in the TCL reaction), may result in enhanced detection sensitivity, perhaps by forming branched structures. Biotin-labeled probes generated by the TCL protocol detect as little as 25 zeptomoles (2.5×10^{-20} moles) of a target sequence. A 50 μ l TCL reaction yields as much as 25 μ g of labeled DNA, enough to probe 25 to 50 Southern blots. After 20 cycles of denaturation and extension, biotin-dUTP-incorporated TCL probes may

- 23 -

be routinely detected at a $1:10^6$ dilution, which is 1000 fold more sensitive than RPL, and indicates that a significant degree of net synthesis or amplification of the probe is occurring. In addition, non-isotopically-labeled TCL probes exhibit a 10-fold improvement in detection sensitivity when compared to RPL-generated probes. ^{32}P -labeled probes generated by the TCL protocol may also detect as little as 50 zeptomoles (2.5×10^{-20} moles) of a target sequence. As little as 10 pg of template DNA is enough to synthesize 5-10 ng of radioactive version of TCL generates probes having extremely high specific activities, e.g. (about 5×10^9 cpm/ μg DNA), which permits 5 to 10-fold lower detection limits than conventional labeling protocols.

There are several advantages to using restriction-generated oligonucleotides for primer extension labeling of DNA. One advantage is the specificity of the primers. All of the oligonucleotides generated by the TCL system are specific for the template utilized, unlike random primer labeling (RPL) which utilizes synthetic oligonucleotides 6-9 bases in length having a random sequence. The amount of primer required for efficient labeling with the TCL system is only 10 ng, compared to the $10 \mu\text{g}$ of random primers utilized for RPL. Due to their short length, random primers anneal very inefficiently above $25-37^\circ\text{C}$, thus RPL is limited to DNA polymerases such as Klenow or T7. The size of the restriction-generated oligonucleotides are longer than the random primers, which extends the hybridization and extension conditions to include a wide variety of temperatures and polymerases. Thus, the use of the restriction-generated sequence-specific oligonucleotides results in more efficient hybridization and extension as compared to RPL. The TCL system has been optimized for labeling with a thermostable DNA polymerase which allows the option of temperature cycling. After 20 cycles of denaturation and extension, a significant amount of amplified TCL probes can be generated. Most importantly, TCL-labeled probes exhibit a 10 fold improvement in detections sensitivity when compared to RPL-generated probes.

- 24 -

Another aspect of the present invention involves a variation of TCL called Universal Thermal Cycle Labelling (UTCL) in which the extension primers are not supplied by CviJI restriction, but rather, are found endogenously in the enzyme preparations of thermostable DNA polymerases. Random sequence DNA is usually co-purified along with the holo-enzyme preparation of the thermostable DNA polymerases, regardless of the source of the enzyme, i.e. native or cloned. However, only the holo-enzyme, and not the exonuclease minus deletion variants, contain the endogenous DNA. Typically, when the holo-enzymes of thermostable polymerases are used in protocols such as the polymerase chain reaction, the presence of such primers can create spurious results. Methods for circumventing the problems of endogenous DNA are described in *PCR Protocols: A Guide to Methods and Applications*, Eds. M. Innis, *et al.*, Academic Press, 1990.

This residual DNA is rather short (approximately 5-25 bases), as assayed by end-labeling with $\gamma^{32}\text{P}[\text{ATP}]$ and polynucleotide kinase and acts as endogenous "random" primers in a TCL-type reaction. UTCL combines the holo-enzyme of a thermostable polymerase from, for example, *Thermus flavus*, with the intact DNA template and is subjected to repeated cycles of denaturation, annealing, and extension. A radioactive- or non-isotopically-labeled deoxynucleotide triphosphate is incorporated during the extension step for subsequent detection purposes. The amplified, labeled probe represents a very heterogeneous mixture of fragments, which appears as a large molecular weight smear when analyzed by agarose gel electrophoresis. Biotin-labeled probes generated by the UTCL protocol detect as little as 25 zeptomoles (2.5×10^{-20} moles) of a target sequence. A 15 μl UTCL reaction yields as much as 5-10 μg of labeled DNA, enough to probe 5 to 10 Southern blots. After 20 cycles of denaturation and extension, biotin-dUTP-incorporated UTCL probes may be routinely detected at a $1:10^6$ dilution, which is 1000 fold more sensitive than RPL, and indicates that a significant degree of net synthesis or amplification of the probe is occurring. In addition, non-isotopically-labeled UTCL probes exhibit

- 25 -

a 10-fold improvement in detection sensitivity when compared to RPL-generated probes. ^{32}P -labeled probes generated by the UTCL protocol may also detect as little as 50 zeptomoles (2.5×10^{-20} moles) of a target sequence. The radioactive version of UTCL generates probes having extremely high specific activities, e.g.
5 (about 5×10^9 cpm/ μg DNA), which permits 5 to 10-fold lower detection limits than conventional labeling protocols.

The present invention is illustrated by the following examples relating to the isolation of a full length viral DNA clone encoding R.CviJI, to the expression of R.CviJI DNA in *E.coli* strain DH5 α F[']MCR and to purification of
10 R.CviJI from this bacterial stain. More particularly, Example 1 provides for the propagation of IL-3A virus and isolation of viral genomic DNA. Example 2 addresses the improved expression of a clone for the viral methylase M.CviJI. Example 3 describes the strategy for isolating and cloning the viral R.CviJI gene by a forced co-cloning strategy of the M.CviJI gene. Example 4 describes the
15 sequencing of cloned IL-3A genomic DNA and identification of the R.CviJI gene. Example 5 relates the methods for purification of CviJI to homogeneity from an *E.coli* strain, DH5 α F[']MCR, transformed with a plasmid which encodes the R.CviJI enzyme. Example 6 details the amino acid sequence analysis of the purified R.CviJI enzyme. Example 7 describes the analysis of CviJI^{*} recognition
20 sequences. Example 8 relates to a technique for producing restriction generated oligonucleotides using CviJI. Example 9 relates the generation of anonymous primers using CviJI. Example 10 describes end-labeling of CviJI restriction generated oligonucleotides. Example 11 describes primer extension labeling of DNA using restriction generated oligonucleotides. Example 12 relates the use of
25 CviJI in thermal cycle labeling of DNA as well as the method of universal thermal cycle labelling. Example 13 provides a method for generation of quasi-random DNA fragments using CviJI. Example 14 describes fractionation of CviJI digested DNA by size using spin column chromatography. Example 15 details the relative cloning efficiency of CviJI digested, size-fractionated DNA by gel elution and

- 26 -

chromatographic methods. Example 16 describes the comparison of cloning efficiency using lambda DNA fragmented by both sonication and CviJI** techniques. Example 17 details the use of CviJI** fragmentation for shotgun cloning and sequencing. Example 18 describes the shotgun cloning of lambda DNA using CviJI. Example 19 describes the use of CviJI in epitope mapping techniques. Example 20 describes the restriction endonuclease reagent CGase I.

Example 1

Propagation of IL-3A Virus

The exsymbiotic *Chlorella-like* alga, NC64A, originally isolated from *Paramecium bursaria* (Karakashian, S.J. and Karakashian, M.W., *Evolution and Symbiosis in the Genus Chlorella and Related Algae. Evolution* 19:368-377 (1965)), was grown and maintained in Bold's basal medium (BBM), (Nichols, H.W. and Bold, H.C. *J. Phycol.* 1:34-38 (1965)) modified by the addition of 0.5% sucrose, 0.1% protease peptone, and 20 μ g/ml tetracycline (MBBM). Cultures were inoculated with 1×10^6 algae cells/ml and grown at 25°C in 250 ml of MBBM in 500 ml Erlenmeyer flasks on a rotary shaker (150 rpm) in continuous light (ca. 30 μ Ei, m⁻²,sec⁻¹). Growth was monitored by light scattering measured as A_{640nm} and/or by direct cell counts with a hemocytometer.

When the cultures reached approximately 1×10^7 algae cells/ml they were inoculated with filter sterilized (0.4 μ m nitrocellulose filter, Nucleopore, Pleasanton, California) IL-3A virus at a multiplicity of infection of 0.01 and incubated for an additional 48 - 72 hours at 25°C. The crude lysate was then centrifuged at 3000 rpm (2000 xg) for 10 minutes to remove cellular debris. Nonidet P-40 was then added to 1% (v/v) and the virus was pelleted from the supernatant by centrifuging at 15,000 rpm at 4°C for 75 minutes in a Beckman No. 30 rotor. The viral pellet was gently resuspended in 0.05 M Tris-HCl, pH

- 27 -

7.8, and the sample was layered on linear 10 - 40% sucrose gradients equilibrated with 0.05 M Tris-HCl, pH 7.8, and centrifuged for 20 minutes at 20,000 rpm at 4°C in a Beckman SW28 rotor. The viral band, which was present in the center of the gradient as an opaque band, was removed, diluted with 0.05 M Tris-HCl, pH 7.8, and pelleted by centrifugation at 15,000 rpm at 4°C for 120 minutes in a Beckman No. 80 rotor. The virus was resuspended in a small volume (10ml) of 0.05 M Tris-HCl, pH 7.8, and stored at 4°C.

IL-3A viral DNA was purified from the viral particles using a modification of the protocol described by (Miller, S.A., Dykes, D.D., and Polesky, H.I., *Nucleic Acids Res.* 16:1215 (1988)). Briefly, 100 µl of IL-3A virus (9.8×10^{11} plaque forming units/ml) was diluted with 400 µl of water and then mixed with 10 µl TEN (0.5 M Tris-HCl, pH 9.0, 20 mM EDTA, 10 mM NaCl) and 10 µl of 10% SDS. After incubating at 70°C for 30 minutes the solution was extracted twice with phenol-chloroform-isoamyl alcohol, extracted once with chloroform, and precipitated with ice-cold ethanol using methods well known in the art and resuspended in 500 µl of H₂O. (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds.) (1987) *Current Protocols in Molecular Biology*, Wiley, New York; Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Example 2

CviJI Methyltransferase Clone

The CviJI methyltransferase gene (M.CviJI) from *Chlorella* virus IL-3A was cloned and sequenced by Shields *et al.*, *Virology* 176:16-24 (1990). Briefly, *Sau3A* partial digest of *Chlorella* virus IL-3A was ligated to *Bam*HI digested pUC19 and transformed into *E. coli* strain RR1. This library of plasmids was restricted with *Hind*III (AAGCTT) and *Sst*I (GAGCTC), both of which are

- 28 -

inhibited by 5-methylcytidine (5mC) in the AGCT portion of their recognition sequences, and transformed again into RR1 cells. M.CviJI methylates the internal cytidine in (G/A)GC(T/C/G) sequences. If the M.CviJI gene is cloned and expressed appropriately, the plasmid DNA would be expected to be resistant to
5 *HindIII* and *SstI* restriction.

The CviJI methyltransferase gene was originally cloned as a 7.2 kb insert, termed pIL-3A.22. Plasmid pIL-3A.22 was only partially resistant to CviJI digestion. Partial digestion is most likely due to the inefficient expression of the M.CviJI gene and the numerous CviJI sites in both the vector (pUC19 has 45
10 CviJI sites) and in the insert DNA. The M.CviJI gene was eventually sublocalized to a region of 3.7 kb by subcloning using methods well known in the art (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds.) (1987) *Current Protocols in Molecular Biology*, Wiley, New York; Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), *Molecular*
15 *Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) and testing the subcloned DNA for sensitivity/resistance to *HindIII*, *SstI*, and CviJI. (Shields *et al.*, *supra*) The entire sequence was determined and three open reading frames which could code for polypeptides 161, 367, and 162 amino acids, respectively, were identified.
20 The 367 amino acid open reading frame (ORF) was identified as the M.CviJI gene by three criteria: (i) it is the only ORF located in the region identified by transposon mutagenesis; (ii) it has amino acid motifs similar to those of other cytosine methyltransferases; and (iii) a 1.6 kb *DraI* fragment containing the 367 amino acid ORF (1101 bp) produces the methyltransferase. This 1.6 kb M.CviJI
25 encoding fragment was subcloned into the *EcoRV* site of pBluescript KS(-) (Stratagene, LaJolla, CA), in the same translational orientation as the *lacZ'* gene of this vector. A physical map of the resulting plasmid termed p710 is shown in Figure 1.

- 29 -

The plasmid p710 was digested with several endonucleases to indirectly test the efficiency of *M.Cvi*II expression. Fully active methylase should render the plasmid DNA completely resistant to digestion by the following enzymes: *Hae*III (which recognizes the sequence GGCC), *Sac*I (which recognizes the sequence GAGCTC), and *Hind*III (which recognizes the sequence AAGCTT). The plasmid was partially resistant to *Hae*III (90%) and *Sac*I (90%), and even less resistant to *Hind*III (25%) digestion. This lack of complete protection of the plasmid DNA made it impractical to attempt cloning the three/two base restriction endonuclease encoded by the *R.Cvi*II gene. Thus, improvements in the efficiency of *M.Cvi*II expression were required before attempting to clone the *R.Cvi*II gene.

The translation efficiency of the *M.Cvi*II gene was improved by removing extraneous 5' open reading frames, creating a perfect fusion of the *lacZ'* Shine-Delgarno sequence with the methyltransferase start codon (see Figure 1). This was achieved by site-specific oligonucleotide mutagenesis, using the oligomer

5'-CAATTTTCACACAGGAAACAGCTATGTCTTTTCGCACGTTAGAAC-3'
(SEQ ID NO: 1) to precisely remove the intervening *lacZ'* DNA. The relevant DNA sequences are indicated in Figure 1 (SEQ ID NO:12). The mutagenesis was facilitated by converting the double stranded plasmid DNA of p710 to single-stranded DNA by co-infecting the *E. coli* host strain with the helper phage R408 (Russel, M., Kidd, S. and Kelly, M.R. *Gene* 45:333-338), using methods well known in the art. The mutagenesis reaction was completed using a commercially available kit according to the manufacturer's instruction (Mutagene, Bio-Rad, Hercules, California). The oligonucleotide was annealed to the single-stranded plasmid, extended in the presence of T4 DNA polymerase, ligated using T4 DNA ligase, and transformed into competent SURE™ cells (Stratagene, La Jolla, California). Transformed cells were then grown overnight as a pool, the DNA isolated and purified.

- 30 -

Enrichment for the mutagenized plasmids was made possible by virtue of the loss of an *Xho*I site located in the sequence that was deleted by mutagenesis. Enrichment was accomplished by digesting the isolated, purified plasmid DNA with *Xho*I, followed by dephosphorylation with calf intestinal alkaline phosphatase (CIAP), and transformed into SURE cells. Plasmid DNA was isolated from 18 individual colonies and the DNA tested for resistance to *Xho*I. Plasmid DNA from 11 colonies were resistant to *Xho*I digestion, indicating that they lacked the deleted sequence. Five of these plasmids were restricted with *Hae*III, *Hind*III, *Pvu*II (which recognizes the sequence CAGCTG), and *Cvi*JI. All five appeared 100% resistant to these enzymes. Four of the plasmids were sequenced and the deletion was confirmed as being correct. One of these, pBMC5, was chosen for further modification.

Example 3

Forced Co-Cloning of R.*Cvi*JI

The location of the R.*Cvi*JI gene on the IL-3A virus genome was inferred as being 3' to the M.*Cvi*JI gene for two reasons: 1) the cloned DNA sequence 5' to the M.*Cvi*JI gene did not produce a restriction activity; and 2) several attempts to clone the DNA 3' to the M.*Cvi*JI gene resulted in deletions/rearrangements of this downstream region. This information permitted a forced co-cloning strategy to obtain the restriction endonuclease gene. This strategy uses a deletion derivative of pBMC5 lacking the 3' half of the M.*Cvi*JI gene. Digestion of the IL-3A genome with the same enzyme used to create the M.*Cvi*JI deletion, followed by ligation of the respective DNAs, transformation, and digestion with enzymes incapable of recognizing methylated DNA (e.g., *Hae*III, *Hind*III, *Pvu*II, *Cvi*JI, etc.) should force the selection of clones which have a restored M.*Cvi*JI gene (and thus active methylase enzyme), as well as downstream DNA. Thus, if a clone is found to be *Cvi*JI resistant, the 3' half of

- 31 -

*M.Cvi*II must have been restored, and downstream DNA containing the *R.Cvi*II gene, at least in part, would presumably be cloned.

The details of this cloning strategy are as follows. pBMC5 has two *Eco*RI sites, one approximately in the middle of the *M.Cvi*II gene, while the other site lies in the vector DNA, 3' to the *M.Cvi*II gene (see Figure 1). pBMC5 was restricted with *Eco*RI and ligated at a dilute concentration (10-50 ng/ μ l) to favor circularization without the 3' *M.Cvi*II fragment. The reaction mixture was then transformed into competent SURE cells and plated on TY agar containing ampicillin. Plasmid DNA from the resulting colonies was tested for the lack of this *Eco*RI fragment by digestion with *Eco*RI. One of these clones, pBMC5RI, was used for the subsequent co-cloning work. Plasmid pBMC5RI was digested with *Eco*RI and dephosphorylated using CIAP. IL-3A genomic DNA was then digested to completion with *Eco*RI. The *Eco*RI digested pBMC5RI and IL-3A DNAs were combined at a ratio of 1:3 in a ligation reaction using T4 DNA ligase, and the products of the ligation reaction were subsequently used to transform competent SURE cells. The pBMC5RI/IL-3A transformants were not plated, but rather grown overnight in culture as a library or pool of cells. The cells were harvested the next day and DNA was isolated and purified. Isolated, purified DNA was digested with *Hae*III, dephosphorylated with CIAP, and transformed into competent SURE cells. The cells were then plated and grown overnight. Six colonies grew, of which only one containing the plasmid, pCJH1.4, was resistant to *Hae*III. The plasmid pCJH1.4 was found to encode *Cvi*II restriction activity. Plasmid pCJH1.4 was further characterized to localize the gene for *Cvi*II by deletion analysis, subcloning experiments, and sequencing. The plasmid pCJH1.4 was deposited with the American Type Culture Collection on June 30, 1993 under Accession Number 69341.

- 32 -

Example 4**Sequencing of Cloned IL-3A DNA Containing CviJI Gene**

5 The *EcoRI* fragment cloned into pCJH1.4 (as described in Example 3) is 4901 bp in length. Except for the 519 bp corresponding to the 3' portion of the *M.CviJI* gene, the remainder of the 4901 bp *EcoRI* fragment cloned into pCJH1.4 was sequenced using the SEQUAL DNA Sequencing System (CHIMERx, Madison, WI) by methods well known in the art. Sequencing was accomplished using three approaches: 1) primer walking on pCJH1.4, 2) cloning various restriction endonuclease digests of pCJH1.4 into an M13 type sequencing vector; and 3) sequencing various restriction endonuclease deletion derivatives of pCJH1.4. The nucleotide sequence of 5497 bp of IL-3A viral DNA is shown in Figure 2 and set forth in SEQ ID NO.: 2.

15 Six open reading frames (ORF) of 1155 bp (ORF1), 468 bp (ORF2), 555 bp (ORF3), 1086 bp (ORF4), 397 bp (ORF5) and 580 bp (ORF6) which could code for polypeptides containing 358 (41.4 kD), 156 (19.4 kD), 185 (20.3 kD), 362 (38.9 kD), 132 (14.5 kD) and 193 (21.9 kD) amino acids, respectively, were identified (see Figure 3). ORFs 4-6 do not code for the *R.CviJI* gene, as the deletion derivative pCdA12, which lacks the DNA between the *AvaI* and *BamHI* sites (see Figure 3), does produce *CviJI* restriction endonuclease activity. In addition, the deletion derivative pCdEB7, lacking the DNA between the *EcoRI* and *BamHI* sites, did not produce *CviJI* activity. Thus ORF1 or ORF3 were the most likely candidates for encoding the *R.CviJI* gene. The sequence of the 1155 bp ORF1 (SEQ ID NO: 3), its deduced amino acid sequence (SEQ ID NO: 4) (as shown in capital letters), plus flanking bases, is presented in Figure 4. The vertical line in Figure 4 and the associated arrow indicate where the DNA sequence from pJCH1.4 diverges from that of pIL-3A.22-8 (Shields, S.L., *et al.*, *Virology* 76:16-24, 1990). This open reading (ORF1) frame is believed to represent the *CviJI* gene because 14 out of 15 N-

- 33 -

terminal amino acids from the protein sequence (see Example 6) matched the predicted translation product of the nucleic acid sequence (Figure 4). Also, the 32.5 kD molecular weight of the homogeneously purified enzyme described in Example 5 matched the predicted translation product of the nucleic acid sequence (31.6 kD) if the encoded protein was translated beginning at the GTG codon located at nucleotides 299 - 301 (Figure 4), instead of the 5' ATG codon located at nucleotides 59 - 61. This possibility is not surprising in light of the fact that approximately 10% of prokaryotic and eukaryotic gene products begin translation with a GTG start codon, rather than the usual ATG codon (Kozak, M., *Microbiol. Rev.* 47:1-45 (1983); Kozak, M. *J. Cell. Biol.* 108:229 (1989); Gold, L. *et al.*, *Annu. Rev. Microbiol.* 35:365-403 (1981)). The structural gene was identified to be 834 nucleotides in length, coding for a protein of 278 amino acids (31.6 kD) and is set forth in SEQ ID NO: 4. It is also interesting to note that the *CviJI* gene was shown to possess an in-frame, upstream ATG codon which if translated could yield a protein with a predicted molecular weight of 41.4 kD (Figure 4). A larger molecular weight form possessing *CviJI* restriction activity has not been detected by SDS gel electrophoresis. However, a second peak of *CviJI* activity which eluted separately from the 32.5 kD form was detected in the initial stages of enzyme purification. The DNA sequence which could theoretically code for a larger form of *CviJI* would be approximately 1074 nucleotides in length (assuming it starts at the upstream ATG codon) and would code for a protein of 358 amino acids.

Example 5

Purification of Recombinant *CviJI* Restriction Endonuclease

Initially, 20 ml of LB medium (plus 100 μ g/ml ampicillin) were inoculated with a 1 ml stock of *E. coli* transformed with the plasmid pCJH1.4 described above and grown overnight at 37°C with shaking. The next day, 20 ml

- 34 -

of this initial overnight culture was used to inoculate another 1 liter of LB medium and grown overnight. The following day, 50 liters of TB medium (12 g Bacto-Tryptone, 24 g Bacto Yeast Extract, 4 ml glycerol, 2.31 g KH_2PO_4 , 12.54 g K_2HPO_4 , 0.1 g MgSO_4 , 100 $\mu\text{g/ml}$ ampicillin, and water to 1 liter) were inoculated with an aliquot of the secondary overnight culture and grown at 37°C with 20 liters/min aeration at 200 RPM, until the $\text{OD}_{595\text{nm}}$ reached 1.0 unit. Vigorous aeration was essential for CviJI expression and a typical yield contained 70 g of cell paste after centrifugation.

The cell pellet was immediately resuspended in lysis buffer A (30 mM Tris-HCl, pH 7.9 at 4°C, 2 mM EDTA, 10 mM beta-mercaptoethanol, 50 $\mu\text{g/ml}$ phenylmethylsulfonyl fluoride (PMSF), 20 $\mu\text{g/ml}$ benzamidine, 2 $\mu\text{g/ml}$ o-phenantroline, 0.7 $\mu\text{g/ml}$ pepstatin) at a volume of 3 ml of buffer A per 1 g of cells. The cell suspension was then passed through a Manton-Gaulin cell disrupter (Gaulin Corporation, Everett, MA) twice and centrifuged for 1 hr (8000 RPM, Sorvall GS3 Rotor) at 4°C. To the supernatant, solid NaCl was added to a final concentration of 200 mM, and 10% polyethyleneimine (PEI) solution slowly added to a final concentration of 1%. The mixture was stirred for 3 hr, and then centrifuged 30 min, at 4°C, 8000 RPM (Sorvall GS3 Rotor). Solid ammonium sulfate was then added to the supernatant at 0.5 g/ml and the mixture was stirred overnight at 4°C. The precipitated proteins were centrifuged for 1 hr. (8000 RPM, Sorvall GS3 Rotor) at 4°C and the resulting pellet dissolved in 100 ml of buffer B (10 mM K/PO_4 , pH 7.2, 0.5 mM EDTA, 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.05% Triton X-100, 50 $\mu\text{g/ml}$ PMFS, 20 $\mu\text{g/ml}$ benzamidine, 2 $\mu\text{g/ml}$ o-phenanthroline, 0.7 $\mu\text{g/ml}$ pepstatin). The dissolved protein solution was then dialysed (14kD cut-off) for 12 hours against three 1 liter changes of buffer B. The dialyzed solution was then diluted to 600 ml with buffer B and applied to a 5 x 20 cm phosphocellulose P11 (Whatman) column (flow rate 100 ml/hr).

- 35 -

The column was then washed with 1.5 liter of buffer B followed by a 0 - 1.5 M NaCl gradient in buffer B (5 liters). R.CviII eluted at approximately 600 mM NaCl. The active fractions were then pooled and concentrated to 50 ml with a 76 mm Amicon YM10 membrane. The resulting solution was then diluted to 300 ml with buffer C (20 mM Tris-acetate, pH 7.4 at 4°C, 2 mM EDTA, 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) and applied to 2.5 x 7 cm Heparin-Sephrose column at a flow rate of 25 ml/hr.

After a 400 ml wash with buffer B, R.CviII was eluted with a 1.5 liter gradient of 0 - 1.3 M NaCl in buffer C. CviII eluted at approximately 400 mM NaCl. The most active fractions were pooled and applied to a 2.5 x 7 cm Blue-agarose column equilibrated in buffer D (20 mM Tris-acetate pH 8.0, 1 mM EDTA, 7 mM beta-mercaptoethanol, 30 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin). After a 500 ml wash with buffer D, CviII was eluted with a 0 - 1.5 M NaCl gradient (1.5 l) in buffer D. Active fractions were dialyzed against buffer G (10 mM K/PO₄ pH 7.0 (4°C), 10 mM beta-mercaptoethanol, 50 mM NaCl, 10% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) and loaded (20 ml/h) onto a ceramic HTP column (American International Chemical, Natick MA) (1.5 x 3 cm), equilibrated in buffer F (20 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 3 mM DTT, 50 mM K-acetate, 5 mM Mg acetate, 50% glycerol). After washing with 100 ml of buffer F, a 400 ml gradient 0 - 0.9 M K/PO₄ in buffer F was run. The HTP column was washed with buffer G, containing 3 mg/ml BSA, then with 1 M phosphate buffer and reequilibrated in buffer G. The active fractions were then pooled and concentrated using a TM10 membrane to a final volume of 3 - 4 ml. This concentrate was then applied to a 2.5 x 95 cm Sephadex G-100 column, equilibrated in buffer E (20 mM Tris-HCl

- 36 -

pH 7.5 (4°C), 5 mM Mg-Acetate, 2 mM EDTA, 10 mM beta-mercaptoethanol, 100 mM NaCl, 5% glycerol, 0.01% Triton X-100, 50 µg/ml PMFS, 20 µg/ml benzamidine, 2 µg/ml o-phenanthroline, 0.7 µg/ml pepstatin) at a flow rate of 6 ml/hr, and 3 ml fractions collected. Active fractions were dialyzed against storage buffer F.

The molecular weight of the purified *Cvi*II was determined by comparison to known protein standards on a denaturing 10% SDS polyacrylamide gel and a single band migrating with an apparent molecular weight of 32.5 kilodaltons was seen indicating that by these criteria, *Cvi*II was purified to homogeneity.

Example 6

N-Terminal Amino Acid Sequence of R.*Cvi*II

To confirm that the restriction endonuclease encoded by the insert in pCJH1.4 was *Cvi*II the sequence of the first 15 N-terminal amino acids of purified *Cvi*II was determined by the Edman degradation method using an Applied Biosystems (Foster City, CA) 477A Liquid Phase Protein Sequencer with an on-line 120A PTH Analyzer. The results of that analysis are shown in Table 1.

- 37 -

Table 1

N-Terminal Amino Acid Analysis of CviJI

	Amino Acid #	Retention Time (min)	pmol (Raw)	Pmol (-bkgd)	Pmol (+lag)	Pmol Ratio	Amino Acid ID
5	1	9.17	6.11	3.86	5.10	34.53	THR, MET, ARG, OR LYS
	2	10.32	3.92	1.54	1.82	9.96	GLU
	3	10.33	4.28	2.22	2.18	11.96	GLU
	4	27.37	2.23	1.49	1.72	7.64	LYS
	5	27.35	2.37	1.66	1.67	7.39	LYS
10	6	17.95	3.37	2.76	2.81	9.48	ARG
	7	28.10	3.19	1.73	2.08	6.09	LEU
	8	13.58	3.58	2.11	2.49	12.08	ALA
	9	28.10	3.23	1.68	1.58	4.63	LEU
	10	18.17	0.71	0.78	0.36	1.21	ILE
15	11	10.30	1.65	0.78	0.96	5.26	GLU
	12	9.72	8.03	0.41	1.31	3.25	LYS
	13	8.53	1.54	0.53	0.55	2.97	GLN
	14	18.18	2.19	1.74	1.67	5.63	ARG
	15	26.80	3.33	0.43	-	0.89	ILE

20 Abbreviations used: threonine (THR), methionine (MET), arginine (ARG), lysine (LYS), glutamic acid (GLU), leucine (LEU), alanine (ALA), isoleucine (ILE) and glutamine (GLN).

The results of this analysis confirm that the protein encoded by the DNA insert in pCJH1.4 (ORF1) is CviJI.

- 38 -

The following Examples illustrate some of the unique properties of and important uses for CviJI.

Example 7

Analysis of CviJI* Recognition Sequences

5 The CviJI* recognition sequence (see Xia, *et al.*, *Nuc.Acids Res.*
15: 6025-6090, 1987) was deduced by cloning and sequencing CviJI* digested
pUC19 DNA fragments. A complete CviJI* digest of pUC19 was ligated to an
M13mp18 cloning derivative for nucleotide sequence analysis. The sequence of
the entire insert was read in order to determine which sites were or were not
10 utilized. A total of 100 clones were sequenced, resulting in 200 CviJI* restricted
junctions, the data for which are compiled in Table 2.

Table 2
Distribution of CwJI* Sites as Assayed by Cloning and Sequencing

<u>Classification Group</u>	<u>NGCN Recognition Sequence</u>	<u>CwJI* Sites Found in pUC19 (%)</u>	<u>CwJI* Sites Cleaved (%)</u>	<u>Sites Not Cleaved (%)</u>	<u>Pu/Py Structure</u>
Normal (N)	A C	AGCC 9 (4.4)	23 (11.5)	1 (0.9)	PuPuPyPy
	GC	GGCC 11 (5.4)	24 (12.0)	1 (0.9)	
	G T	GGCT 10 (4.9)	13 (6.5)	0 (0.0)	
		AGCT 15 (7.3)	35 (17.5)	0 (0.0)	
		45 (22.0)	95 (47.5)	2 (1.7)	
Relaxed (R1)	C C	CGCC 11 (5.4)	11 (5.5)	4 (3.5)	PyPuPyPy
	GC	TGCC 12 (5.9)	13 (6.5)	10 (8.6)	
	T T	TGCT 10 (4.9)	10 (5.0)	5 (4.3)	
		CGCT 22 (10.7)	17 (8.5)	7 (6.0)	
		55 (26.0)	51 (25.5)	26 (22.4)	
Relaxed (R2)	A A	AGCA 16 (7.3)	13 (6.5)	5 (4.3)	PuPuPuPu
	GC	GGCA 8 (3.9)	11 (5.5)	3 (2.6)	
	G G	AGCG 11 (5.4)	12 (6.0)	11 (9.5)	
		GGCG 22 (10.7)	18 (9.0)	8 (6.9)	
		57 (27.8)	54 (27.0)	27 (23.3)	
Relaxed (R3)	C A	CGCA 10 (4.9)	0	12 (10.4)	PyPuPyPu
	GC	TGCA 13 (6.3)	0	19 (16.4)	
	T G	CGCG 10 (4.9)	0	27 (23.3)	
		TGCG 15 (7.3)	0	3 (2.6)	
		48 (23.4)	0	61 (51.6)	
		Total 205	200	116	

- 40 -

The dinucleotide GC is found at 205 sites in pUC19. These GC sites (shown in Table 2) can be divided into four classes based on their flanking Pu/Py structure, the normal recognition sequence (N) and three potential classes of relaxed sites (R2 and R3). As seen in Table 2, the fraction of such NGCN sites which belong to each classification is roughly equal (22.0%-27.8%). A total of 200 *Cvi*JI* restricted junctions were analyzed by sequencing 100 cloned inserts. If *Cvi*JI* cleaved at all NGCN sites without sequence preferences, it would be expected that the fraction of each classification should be restricted approximately equally. Instead, most of the sites cleaved by this treatment were found to be normal, or PuGCPy sites (47.5%). R1 (PyGCPy) and R2 (PuGCPu) restricted sites were found at nearly the same frequency (25.5% and 27.0%, respectively). Out of 200 *Cvi*JI* junctions, no R3 (PyGCPu) restricted sites were found. Thus, *Cvi*JI* cleaves all NGCN sites except for PyGCPu. As *Cvi*JI* cleaves 12 out of 16 possible NGCN sites, it may be referred to as a 2.25-base recognition endonuclease.

In addition to the restricted sites, those sites which were not cleaved by *Cvi*JI* conditions were also compiled for analysis, as shown in Table 2. A total of 116 non-cleaved NGCN sites were found in the 100 inserts which were sequenced. PyGCPu sites represented the largest class of non-cleaved sites (52.6%). In only two cases were PuGCPy sites found not to be cleaved. An approximately equal fraction of R1 and R2 sites were not cleaved as were found cleaved (22.4% versus 25.5% for R1 and 23.3% versus 27.0% for R2). Based on the frequency of cleavage, or lack thereof, a hierarchy of restriction under *Cvi*JI* conditions is evident, where PuGCPy >> PuGCPu = PyGCPy.

- 41 -

Example 8

CviJI* Restriction Generated Oligonucleotides

Due to the high frequency of CviJI or CviJI* restriction, it is possible to generate useful oligonucleotides by digestion and a heat denaturation step as described above. The size and number of the resulting oligonucleotides are important for subsequent applications such as those described above. If for example, an oligonucleotide is to be used with a large genome, it has to be long enough so that the sequence detected has a probability of occurring only once in the genome. This minimum length has been calculated to be 17 nucleotides for the human genome (Thomas, C.A., Jr. *Prog. Nucl. Acid Res. Mol. Biol.*, 5:315 (1966)). Oligonucleotides used for sequencing or PCR amplification are generally 17-24 bases in length. Oligomers of shorter length will often bind at multiple positions, even with small genomes, and thus will generate spurious extension products. Thus, an enzymatic method for generating oligomers should ideally result in polymers greater than 18 bases in length.

The theoretical number of pUC19 CviJI* restriction-generated oligomers is 314 (157 CviJI* restriction fragments x 2 oligomers/fragment), the size distribution of which is shown in panel A of Figure 5. Most of the expected CviJI* restriction-generated oligomers (about 75%) are smaller than 20 bp. This assumes that CviJI is capable of restricting DNA to very small fragments, the shortest of which would be 2 bp. However, in practice, about 93% of the cloned CviJI* fragments were 20-56 bp in size, and 3% of the fragments generated by CviJI* were smaller than 20 bp (panel B of Figure 5). This suggests that CviJI* is not able to bind or restrict those fragments below a certain threshold length. Since the smallest observed fragment is 18 bp, it may be assumed that this length is the minimal size which can be generated from a given larger fragment. Whatever the reason for this phenomenon, CviJI* treatment of DNA produces a

- 42 -

relatively small range of oligomers (mostly 20-60 bases in length), most of which are a perfect size class for molecular biology applications.

Example 9

Anonymous Primer Cloning

5 Primers are critical tools in many molecular biology applications such as PCR, sequencing, and as probes. Anonymous primers are useful as sequencing primers for genomic sequencing projects, as probes for mapping chromosomes, or to generate oligonucleotides for PCR amplification.

10 The Anonymous Primer Cloning (APC) method is a variation of shotgun cloning in that unknown sequences of DNA are being randomly cloned. However, unlike *Cvi*JI shotgun cloning, wherein a partial *Cvi*JI^{**} digest of DNA is cloned, anonymous primer cloning utilizes a complete *Cvi*JI^{*} digest to restrict large DNAs into small fragments 20-200 bp in size. These small fragments are cloned into a unique vector designed for excising the anonymous DNA as labeled primers. The strategy for this method is illustrated in Figure 6.

15 As illustrated in Figure 6, the APC strategy reduces large DNAs to small fragments, which are cloned and excised for use as primers. Plasmid pFEM has a unique arrangement of the restriction sites for *Mbo*II and *Fok*I, which permits DNA cloned into the *Eco*RV site to be excised without associated vector DNA. This is possible because *Fok*I cleaves 9/13 bases to the left of the recognition site shown in pFEM and *Mbo*II cleaves 8/7 bases to the right of the recognition site shown in pFEM, which is well into the cloned anonymous sequence. After *Mbo*II or *Fok*I restriction, a known flanking primer is annealed (primer 1 or 2) and extended using a DNA polymerase and dNTPs. The primer
20 is previously end-labeled, or alternatively, one or more
25 radioactive.

- 43 -

After denaturation of the newly synthesized DNA and separation from its cognate template, the labeled anonymous primer is ready for use in sequencing the original template from which it was subcloned. The presence of the pFEM vector sequence fused to the anonymous sequence does not influence the enzymatic extension of this primer from its unique binding site, as the vector DNA is at the 5' end and the unique sequence is located at the 3' end (all polymerases extend 5' to 3'). Both the top and bottom strand primers may be excised from pFEM due to the symmetrical placement of restriction sites and flanking primer binding sites. Thus, two primers may be derived from each cloning event. APC is particularly well suited to the genomic sequencing strategy of Church and Gilbert *Proc Natl. Acad Sci. USA* **81**:1991-1995 (1984), although its utility is not limited thereto.

Example 10

End Labeling of Restriction-Generated Oligonucleotides

As is clear from the foregoing examples, digesting DNA with CviII* provides the ability to generate sequence-specific oligonucleotides ranging in size from 20-200 bases in length with an average length of 20-60 bases. Sequence specific oligonucleotides generated by CviII* digestion may be labeled directly at the 5'-end or at the 3'-end using techniques well known in that art.

For example, 5'-end labeling may be accomplished by either a forward reaction or an exchange reaction using the enzyme T4 polynucleotide kinase. In the forward reaction, ^{32}P from [$\gamma\text{-}^{32}\text{P}$]ATP is added to a 5' end of an oligonucleotide which has been dephosphorylated with alkaline phosphatase using standard techniques widely known in the art and described in detail in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition. Cold Spring Harbor Laboratory Press (1989). In an exchange reaction, an excess of ADP (adenosine diphosphate) is used to drive an exchange of a 5'-terminal phosphate

- 44 -

from the sequence specific oligonucleotide to ADP which is followed by the transfer of ^{32}P from $\gamma^{32}\text{P}$ -ATP to the 5'-end of the oligonucleotide. This reaction is also catalyzed by T4 polynucleotide kinase and is described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition*. Cold Spring Harbor Laboratory Press (1989).

Homopolymeric tailing is another standard labeling technique useful in the labeling of CviJI^* -generated sequence specific oligonucleotides. This reaction involves the addition of ^{32}P -labeled nucleotides to the 3'-end of the sequence specific oligonucleotides using a terminal deoxynucleotide transferase. (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Edition*. Cold Spring Harbor Laboratory Press (1989)).

Commonly used labeling techniques typically employ a single oligonucleotide directed to a single site on the target DNA and containing one or a few labels. Oligonucleotides generated by the method of the present invention are directed to many sites of a target DNA by virtue of the fact that they are generated from a sample of the target sequence. Thus, the hybridization of multiple oligonucleotides (labeled by the methods described above) allows a significantly enhanced sensitivity in the detection of target sequences. In addition, the short length of the labeled oligonucleotides used in the methods of the present invention allows a reduction in hybridization time from overnight (as is used in conventional methods) to 60 mins.

Although labeling sequence specific oligonucleotides with ^{32}P is described above, labeling with other radionucleotides, and non-radioactive labels is also within the scope of the present invention.

- 45 -

Example 11**Primer Extension Labeling of DNA Using
Restriction-Generated Oligonucleotides (PEL-RGO)**

Another aspect of the present invention includes methods for labeling DNA which include the generation of oligonucleotide primers by complete digestion with CviJI*, followed by heat denaturation. PEL-RGO requires three steps: 1) generating the sequence-specific oligonucleotides by CviJI* restriction of the template DNA; 2) denaturation of the template and primer; and 3) primer extension in the presence of labeled nucleotide triphosphates. Plasmid DNA may be prepared by methods known in the art such as the alkaline lysis or rapid boiling methods (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)). In addition, the vector should be linearized to ensure effective denaturation. A restriction fragment may be labeled after separation on low melting point agarose gels by methods well known in the art.

In PEL-RGO labeling, template DNA to be labeled is divided into two aliquots; one is used to generate the sequence specific oligonucleotide primers and the other aliquot is saved for the primer annealing and extension reaction. A typical reaction mix for generating sequence-specific oligonucleotides is assembled in a microcentrifuge tube and includes: 100 ng DNA; 2 μ l 5x CviJI* buffer; 0.5 μ l CviJI (1u/ μ l); sterile distilled water to 10 μ l final volume. CviJI* 5X restriction buffer includes: 100 mM glycylglycine (Sigma, St. Louis, Missouri, Cat. No. G2265) pH adjusted to 8.5 with KOH, 50 mM magnesium acetate (Amresco, Solon, Ohio, Cat. No. P0013119), 35 mM β -mercaptoethanol (Mallinckrodt, Paris, Kentucky, Cat. No. 60-24-2), 5 mM ATP, 100 mM dithiothreitol (Sigma, St. Louis, Missouri, Cat. No. D9779) and 25% v/v DMSO, (Mallinckrodt Cat. No. 67-68-5). CviJI is obtained from CHIMERx (Madison, Wisconsin). The reaction mix is incubated at 37°C for 30 min, followed by the

- 46 -

inactivation of *Cvi*II by heating at 65°C for 10 min. The *Cvi*II*-restricted DNA may be used directly without further purification, or it may be stored at -20°C for several months for subsequent labeling reactions.

5 After heat-inactivating *Cvi*II, 0.2 µg of the digested and undigested DNA are electrophoresed on a 1.5% agarose gel, using a suitable molecular weight marker for comparison. The *Cvi*II restriction fragments appear as a low molecular weight smear in the 20-200 bp range.

10 By way of example, 1-10 ng of linearized pUC19 was labeled under the conditions described below. A template-primer cocktail was prepared by mixing 10 ng of linearized pUC19 DNA template with 20 ng pUC19 sequence-specific oligonucleotides (prepared as described above) and the mixture is brought to a final volume of 17 µl with sterile distilled water. The template-primer mixture is denatured in a boiling water bath for 2 minutes and immediately placed on ice.

15 The following labeling mixture is then added to the template-primer mix: 2.5 µl 10X labeling buffer (500 mM Tris HCl at pH 9.0, 30 mM MgCl₂, 200 mM (NH₄)₂SO₄, 20µM dATP, 20µM dTTP, 20µM dGTP, 0.4% NP-40); 5.0 µl [α -³²P] dCTP (3000Ci/mmol, 10µCi/µl New England Nuclear, Catalog No. NEG013H); 0.5 µl *Thermus flavus* DNA polymerase (5u/µl) (Molecular Biology Resources, Milwaukee, Wisconsin); up to 25 µl final volume with distilled water. The reaction was incubated at 70°C for 30 min and then stopped by adding 2µl of 0.5M EDTA at pH 8.0 to the reaction mix.

20 The efficiency of the labeling reaction is gauged by the percentage of radioisotope incorporated into labeled DNA. One microliter of the labeling reaction is added to 99 µl of 10mM EDTA in a microcentrifuge tube. This serves as the source of diluted probe for total and trichloroacetic acid (TCA)-precipitable counts. 2 µl of diluted probe is spotted onto the center of a glass fiber filter disc (Whatman number 934-AH). The disc is then allowed to dry and is then placed in a vial containing scintillation cocktail for counting total radioactivity in a liquid

- 47 -

scintillation counter. Another 2 μ l aliquot from the diluted probe is added to 1 ml of 10% ice cold TCA followed by the addition of 2 μ l of carrier bovine serum albumin (BSA). This mixture was then placed on ice for 10 minutes. The precipitate is then collected on a glass filter disc (Whatman No. 934-AH) by vacuum filtration. The filter is then washed with 20ml of ice cold 10% TCA, allowed to dry and is placed in a vial containing scintillation cocktail and counted.

Because primer extension oligonucleotide labeling results in net DNA synthesis, the specific activity of labeled DNA is calculated using the following guidelines.

$$\text{Total cpm incorporated} = \text{TCA cpm} \times 50 \times 27$$

Wherein the factor 50 is derived from using 2 μ l of a 1:100 dilution for TCA precipitation. The number 27 converts this back to the total reaction volume (which is the reaction volume plus 2 μ l of stop solution).

$$\text{Synthesized DNA (ng of DNA synthesized)} = \text{theoretical yield} \times \text{fraction of radioactivity incorporated.}$$

$$\text{Theoretical yield (ng of DNA)} = \frac{\mu\text{Ci dNTPs added} \times 4 \times 330\text{ng/nmole}}{\text{specific activity dNTP (Ci/nmole)} = \mu\text{Ci/nmole}}$$

$$\text{Fraction of incorporated label} = \text{TCA precipitated cpm} / \text{total cpm.}$$

$$\text{Specific activity (cpm}/\mu\text{g of DNA)} = \frac{\text{total cpm incorporated} \times 1000}{\text{synthesized DNA} + \text{input DNA}}$$

Wherein 1000 is the factor converting nanograms to micrograms.

- 48 -

By way of example, the following represents the calculation of specific activity for an aliquot of pUC19 DNA labeled using this method. Using 50 μCi of [α - ^{32}P]dCTP in a 25 μl reaction, and if the TCA precipitated cpm is 26192 and total cpm is 102047;

$$5 \quad \text{Total cpm incorporated} = 26192 \times 50 \times 27 = 3.27 \times 10^7 \text{ cpm}$$

$$\text{Synthesized DNA (ng of DNA synthesized)} =$$

$$\text{Theoretical yield} \times \text{fraction of radioactivity incorporated.}$$

$$\text{Theoretical yield} = \frac{\mu\text{Ci of dNTPs} \times 4 \times 330}{3000 \mu\text{Ci/nmole}}$$

$$10 \quad = \frac{50 \mu\text{Ci} \times 4 \times 330}{3000}$$

$$= 22 \text{ ng}$$

$$\text{Fraction of label incorporated} = \frac{\text{TCA precipitated cpm}}{\text{Total cpm}} = \frac{26192}{102047} = 0.256$$

$$15 \quad \text{Synthesized DNA} = 22 \times 0.256 \\ = 5.6 \text{ ng}$$

$$\text{Specific activity (cpm / } \mu\text{g)} = \frac{\text{Total cpm incorporated}}{\text{Synthesized DNA} + \text{input DNA}} \times 1000$$

$$\text{Input DNA} = 10 \text{ ng}$$

$$20 \quad \text{Specific activity} = \frac{3.27 \times 10^7 \times 1000}{5.6 + 10} \\ = 2.09 \times 10^9 \text{ cpm / } \mu\text{g}$$

Unincorporated radioactive label may be removed using standard methods well known in the art.

- 49 -

Comparisons were made between PEL-RGO vs RPL under similar conditions, and it was observed that a detection limit of 100 fg was seen using PEL-RGO labeled DNA compared to a detection limit of 500 fg with RPL, using a radiolabeled probe.

5

Example 12**Thermal Cycle Labeling and Universal Thermal Cycle Labeling**

Thermal Cycle Labeling (TCL) is a method according to the present invention for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of the labeled probe. TCL of DNA requires two
10 general steps: 1) generation of the sequence-specific oligonucleotides by *Cvi*II* restriction of the template DNA; and 2) repeated cycles of denaturation, annealing, and extension in the presence of a thermostable DNA polymerase or a functional fragment thereof which maintains polymerase activity. Optimal results are obtained after 20 such cycles, which is best performed in an automated
15 thermal cycling instrument such as a Perkin-Elmer Model 480 thermocycler. In conjunction with such an instrument, about 1.5 hr. is required to complete this protocol. If a thermal cycler is not available these reactions may be performed using heat blocks. As few as 5 cycles may yield probes with acceptable detection sensitivities. The generation of sequence specific oligonucleotides for use in this
20 method may also be accomplished using the restriction endonuclease reagent CGase I described in Example 20 or the restriction endonuclease *Aci* I which has as a recognition sequence CCGC.

Non-radioactive labeling of DNA using TCL is accomplished by mixing: 10 pg - 100 ng linearized template, 50 ng *Cvi*II*-digested primers
25 (prepared as described above), 1.5 μ l 10X labeling buffer, 0.5 μ l *Thermus flavus* DNA polymerase (5u/ μ l) (Molecular Biology Resources, Inc., Milwaukee,

- 50 -

Wisconsin), 1 μ l of 1mM Biotin-11-dUTP (Enzo Diagnostics, New York, New York), 1.5 μ l each of dATP, dCTP, and dGTP (2 mM), and 1.0 μ l 2mM dTTP.

Radioactive labeling of DNA using TCL was accomplished by mixing 10 pg - 100 ng of CviJI generated primers, 10 pg-25 ng of linearized
5 template, 1.5 μ l of 10X labeling buffer, 5 μ l of 32 P-dCTP (3000 Ci/mmol, 10 μ Ci/ μ l or 40 μ Ci/ μ l), 0.5 μ l of *Thermus flavus* DNA polymerase (5u/ μ l), and 0.5 μ l each of dATP, dGTP, and dTTP (1 mM) was added. The reaction mix was brought to a volume of 15 μ l with deionized H₂O, overlaid with mineral oil and cycled through 20 rounds of denaturation, annealing and extension. A typical
10 cycling regimen employed 20 cycles of denaturation at 91°C for 5 sec, annealing at 50°C for 5 sec and extension at 72°C for 30 sec. The reaction is then terminated by adding 1 μ l of 0.5M EDTA, pH 8.0. The amplified, labeled probe is a very heterogeneous mixture of fragments, which appears as a smear when analyzed by agarose gel electrophoresis.

15 Universal thermal cycle labeling (UTCL) is a method according to the present invention for efficiently labeling double-stranded DNA while simultaneously amplifying large amounts of labeled probe. UTCL is unique in that no sequence information is required regarding the template. The extension primers are supplied endogenously via the holo-enzyme of the thermostable DNA
20 polymerase and any anonymous DNA template can be labeled by repeated cycles of denaturation, annealing, and extension in the presence of a labeled deoxynucleotide triphosphate. Optimal results are obtained after 20 such cycles, which is best performed in an automated thermal cycling instrument such as a Perkin-Elmer Model 480 thermocycler. In conjunction with such an instrument,
25 about 1.5 hr are required to complete this protocol. If a thermal cycler is not available these reactions may be performed using heat blocks. As a few as 5 cycles may yield probes with acceptable detection sensitivities.

Non-radioactive labeling of DNA using UTCL is accomplished by mixing: 10 ng linearized template, 1.5 μ l 10X labeling buffer, 0.5 μ l *Thermus*

- 51 -

flavus DNA polymerase (5u/ μ l) (Molecular Biology Resources, Inc., Milwaukee, Wisconsin), 1 μ l of 1mM Biotin-11-dUTP (Enzo Diagnostics, New York, New York), 1.5 μ l each of dATP, dCTP, and dGTP (2 mM), and 1.0 μ l 2mM dTTP.

Radioactive labeling of DNA using UTCL was accomplished by
5 mixing: 10 pg-100 ng of linearized template, 1.5 μ l of 10X labeling buffer, 5 μ l of 32 P-dCTP (3000 Ci/mmol, 10 μ Ci/ μ l or 40 μ Ci/ μ l), 0.5 μ l of *Thermus flavus* DNA polymerase (5u/ μ l), and 0.5 μ l each of dATP, dGTP, and dTTP (1 mM) was added. The reaction mix was brought to a volume of 15 μ l with deionized H₂O, overlaid with mineral oil and cycled through 20 rounds of denaturation,
10 annealing and extension. A typical cycling regimen employed 20 cycles of denaturation at 91°C for 5 sec, annealing at 50°C for 5 sec and extension at 72°C for 30 sec. The reaction is then terminated by adding 1 μ l of 0.5M EDTA, pH 8.0. The amplified, labeled probe is a very heterogeneous mixture of fragments, which appears as a smear when analyzed by agarose gel electrophoresis.

15

Estimation of Bio-11 dUTP incorporation:

In order to estimate the level of incorporation of biotin-11-dUTP into DNA, a serial dilution from 1:10 to 1:10⁸ of the labeled probe (free of unincorporated biotin-11-dUTP) is made in TE (10mM Tris, 1mM EDTA, pH 8). A microliter of each dilution is placed on a neutral nylon membrane, and the
20 DNA sample is bound to the membrane either by UV cross linking for 3 min or by baking at 80°C for 2 hr.

The unbound sites on the membrane are blocked using a blocking buffer for 15 min at 25°C. Streptavidin-alkaline phosphatase (Gibco-BRL Gaithersburg, Maryland, Cat. No. 9545A) is added to the blocking buffer (0.058
25 M Na₂HPO₄, 0.017 M NaH₂PO₄, 0.068 M NaCl, 0.02% sodium azide, 0.5% casein hydrolysate, 0.1% Tween-20) at a 1:5000 dilution and incubated for a 30 min., and the membrane is rinsed 3 times for 10 min. each with wash buffer (1x PBS [0.058 M Na₂HPO₄, 0.017 M NaH₂PO₄, 0.068 M NaCl], 0.3% Tween,

- 52 -

0.2% sodium azide), rinsed briefly (5 minutes) with AP buffer (100 mM NaCl, 5 mM MgCl₂, 100 mM Tris-Cl pH 9.5) and then enough AP buffer containing 4.0 µl/ml nitro blue tetrazolium (NBT) (Sigma Cat. No. N6639), (Sigma Cat. No. B6777), and 3.5 µl/ml of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) was added
5 in order to cover the membrane. The membrane is left in the dark for approximately 30 minutes or until the reaction is complete. The reaction is stopped by rinsing in 1 X PBS.

Detection Sensitivities

32P-labeled probes generated by the protocol above described
10 labelling detect as little as 25 zeptomoles (2.5×10^{-20} moles) of a target sequence. As little as 10 pg of template DNA is enough to synthesize 5-10 ng of radiolabeled probe, which is sufficient for screening 5 Southern blots. The radioactive versions of TCL and UTCL facilitate extremely high specific activities of labeled probe (about 5×10^9 cpm/µg DNA), which permits 5-10 fold lower
15 detection limits than conventional labeling protocols. The synthesis of higher specific activity probes is probably the net result of the sequence-specific oligonucleotide primers and their increased length when compared to the short random primers used in other labeling methods. In addition, the thermal cycling permits probe amplification.

20 Biotin-labeled probes generated by the TCL and UTCL protocols detect as little as 25 zeptomoles (2.5×10^{-20} moles) of a target sequence. A 15 µl TCL or UTCL reaction yields as much as 5-10 µg of labeled DNA, enough to probe 5 to 10 Southern blots. Biotin-labeled TCL and UTCL probes provide a 10 fold greater detection sensitivity when compared to RPL biotin probes. In
25 addition, the thermal cycling permits probe amplification.

Non-radioactive, biotinylated probes labeled by the TCL and UTCL methods were shown to have detection limits that are identical to the radioactive probes. These methods have the advantage of eliminating the need to work with

- 53 -

hazardous radioactive materials without sacrificing sensitivity. In addition, results are obtained from non-isotopic probes in 3-4 hours compared to 3-4 days for radiolabeled probes. The ability to substitute non-radioactive probes for radioactive probes may be very useful to clinical laboratories, which do not use radioisotopes but do need greater detection sensitivities. Research laboratories favor the use of non-isotopic systems if detection sensitivity is not an issue. The non-isotopic labeling version of the TCL and UTCL systems represent a major improvement in labeling DNA probes. Non-radioactive probes generated by the methods of the present invention are also useful in the detection of RNA *in situ*. An advantage of this system is that labeling protocols of the present invention yield highly sensitive non-radioactive probes, and the size of the probes are predominantly in the small molecular weight range and can therefore penetrate the tissue easily, unlike RPL. Because non-radioactive probes labeled using the labeling protocols of the present invention have the same detection limits as do radioactive probes similarly labeled, it is within the scope of this invention to use either radioactive or non-radioactive probes for probing, for example, Southern blots, Northern blots, for *in situ* hybridization for the detection of mRNA or DNA in cells or tissue directly, and for colony or plaque lifts.

Example 13

Quasi-Random Fragmentation of DNA

Shotgun cloning and sequencing requires the generation of an overlapping population of DNA fragments. Therefore, conditions were established for the partial digestion of DNA with CviJI to produce an apparently random pattern, or smear, of fragments in the appropriate size range. Conventional methods for obtaining partially restricted DNA include limiting the incubation time or limiting the amount of enzyme used in the digestion. Initially,

- 54 -

agarose gel electrophoresis and ethidium bromide staining of the treated DNA were utilized to assess the randomness and size distribution of the fragments.

CviJI was obtained from CHIMERx (Madison, Wisconsin). Digestion of pUC19 DNA for limited time periods, or with limiting amounts of CviJI under normal or relaxed conditions, did not produce a quasi-random restriction pattern, or smear. Instead, a number of discrete bands were observed, as shown in Figure 7, lane 3 for the CviJI* partial digestion of pUC19. Complete digests of pUC19 under normal and CviJI* buffer conditions are shown in lanes 1 and 2 respectively. These results show that, under these relaxed conditions, CviJI has a strong restriction site preference.

To eliminate the apparent restriction site preferences observed under the partial restriction conditions described above, a series of altered reaction conditions were explored. Conditions of high pH, low ionic strength, addition of solvents such as glycerol or dimethylsulfoxide, and/or substitution of Mn^{2+} for Mg^{2+} were systematically tested with CviJI endonuclease using the plasmid pUC19. Figure 7 shows the results of these tests. In Lane M, a 100 bp DNA ladder was run. In Lanes 1-4, pUC19 DNA (1.0 μ g) was run after digestion at 37°C in a 20 μ l volume for the following times and conditions: Lane 1, complete CviJI digest (1 unit of enzyme for 90 min in 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl); Lane 2, complete CviJI* digest (1 unit of enzyme for 90 min in 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM ATP, 20 mM DTT); Lane 3, partial CviJI* digest (0.25 units of enzyme for 30 min in 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM ATP, 20 mM DTT); Lane 4, partial CviJI** digest (0.5 units of enzyme for 60 min in 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM NaCl, 1 mM ATP, 20 mM DTT, 20% v/v DMSO); and Lane 5, uncut pUC19 (1.0 μ g).

The digestion condition which yielded the best "smearing" pattern was obtained when the ionic strength of the relaxed reaction buffer was lowered and an organic solvent was added (Figure 7, lane 4). Plasmid pUC19 partially

- 55 -

digested under these conditions yields a relatively non-discrete smear. This activity is referred to as CviJI** to differentiate it from the originally-characterized star activity described in Xia *et al.*, *Nucl. Acids Res.* **15**:6075-6090 (1987). The appearance of diffuse, faint bands overlying a background smear
5 generated from this 2686 bp molecule indicates that some weakly preferred or resistant restriction sites may bias the results of subsequent cloning experiments.

DNA was mechanically sheared by sonication utilizing a Heat Systems Ultrasonics (Farmingdale, New York) W-375 cup horn sonicator as specified by Bankier *et al.*, *Methods in Enzymology* **155**:51-93 (1987). DNA
10 fragmented by this method has random single-stranded overhanging ends (ragged ends).

CviJI* digested, and sonicated samples were size fractionated by agarose gel electrophoresis and electroelution, or by spin columns packed with the size exclusion gel matrix, Sephacryl S-500 (Pharmacia LKB, Piscataway N.J.) to
15 eliminate small DNA fragments. Spin columns (0.4 cm in diameter) were packed to a height of 1.3 cm by adding 1 ml of Sephacryl S-500 slurry and centrifuging at 2000 RPM for 5 minutes in a Beckman CPR centrifuge. The columns were rinsed 3 times with 1 ml aliquots of 100 mM Tris-HCl (pH 8.0) by centrifugation at 2000 RPM for 2 min. Typically, 0.2-2.0 µg of fragmented DNA in a total
20 volume of 30 µl was applied to the column. The void volume, containing those DNA fragments larger than 500 bp, was recovered in the column eluant after spinning at 2000 RPM for 5 minutes. The capacity of this micro-column procedure is 2 µg of DNA. Agarose gel electrophoresis and electroelution are described in detail by Sambrook *et al.* *Molecular Cloning: A Laboratory Manual*,
25 Second Edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor N. Y. (1989) and is well known to those skilled in the art. In these experiments, 5 µg of sample was pipetted into a 2 cm-wide slot on a 1% agarose gel. Electrophoresis was halted after the bromophenol blue tracking dye had migrated

- 56 -

6 cm. Fragments larger than 750 bp, as judged by molecular size markers, were separated from smaller sizes and electrophoresed onto dialysis tubing (1000 MW cutoff). The fractionated material was extracted with phenol-chloroform and precipitated using ice cold ethanol (50% final volume) and ammonium acetate (2.5 M final concentration).

The ragged ends of the sonicated DNA were rendered blunt utilizing two different end repair reactions. In one end repair reaction (ER 1) sonicated DNA was treated according to the procedure outlined by Bankier *et al. Methods in Enzymology* 155:51-93 (1987), where 2.0 μ g of sonicated lambda DNA is combined with 10 units of the Klenow fragment of DNA polymerase I, 10 units T4 DNA polymerase, 0.1 mM dNTPs, (deoxynucleotide triphosphates=deoxyadenosine triphosphate, deoxythymidine triphosphate, deoxycytosine triphosphate, and deoxyguanosine triphosphate) and reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT). This mixture was incubated at room temperature for 30 min followed by heat denaturation of the enzymes at 65°C for 15 minutes. In a second end repair reaction (ER 2), an excess of the reagents and enzymes described above were utilized to ensure a more efficient conversion to blunt ends. In this reaction, 0.2 μ g of the sonicated lambda DNA sample was treated under the same reaction conditions described above.

Figure 8 shows comparisons of the size distributions of sonicated DNA versus DNA that was partially digested with *Cvi*JI^{**}. In Lanes M, a 1 kb DNA ladder was run. In Lanes 1-3, untreated λ DNA (0.25 μ g), sonicated λ DNA (1.0 μ g), and *Cvi*JI^{**} partially-digested λ DNA (1.0 μ g) were run, respectively. In Lanes 4-6, untreated pUC19 (0.25 μ g), sonicated pUC19 (1.0 μ g), and *Cvi*JI^{**} partially-digested pUC19 (1.0 μ g) were run, respectively.

Fragmentation of a large substrate such as lambda DNA (45 kb) revealed essentially no banding differences between the *Cvi*JI^{**} method and sonication, as demonstrated in Figure 8, lanes 2 and 3. In addition, pUC19 DNA

- 57 -

that was partially digested with *Cvi*JI** gave a size distribution or "smear" that closely resembled that achieved with sonication (Figure 8, lanes 5 and 6). As expected, the minor bias evident with a small molecule such as pUC19 was not detectable with a larger substrate such as lambda DNA.

5 The intensity and duration of sonic treatment affects the size distribution of the resulting DNA fragments. The results obtained from the sonication of lambda and pUC19 samples (Figure 8) were obtained from three 20 second pulses at a power setting of 60 watts. Sonication-generated smears are similar, although the size distribution of fragments is consistently greater with
10 *Cvi*JI** fragmentation. This result favors the cloning of larger inserts, which facilitates the efficiency of end-closure strategies (Edwards *et al.*, *Genome* 6:593-608 (1990)). The size distribution of the DNA fragmented by *Cvi*JI** is controlled by incubation time and amount of enzyme, variables which are readily optimized by routine analysis. An excess of enzyme or a long incubation time
15 will completely digest pUC19 DNA, resulting in fragments which range in size from approximately 20 bp to approximately 150 bp (Figure 7, lanes 1 and 2). The results shown in Figure 8 were obtained by incubating pUC19 for 40 minutes and lambda DNA for 60 minutes with 0.33 units of *Cvi*JI/ μ g substrate. The efficiencies of the two methods for randomly fragmenting DNA were
20 quantitatively analyzed for use in molecular cloning, as described below.

Example 14

Rapid DNA Size Fractionation Utilizing Spin Column Chromatography

25 The amount of data obtained by the shotgun sequencing approach is substantially increased if fragments of less than 500 bp are eliminated prior to the cloning step. Small fragments yield only a portion of the sequence data which may be collected from polyacrylamide gel based separations and, thus, such small fragments lower the efficiency of this strategy. Agarose gel electrophoresis

- 58 -

followed by electroelution is commonly used to size fractionate DNA prior to shotgun cloning (Bankier *et al.*, *Methods in Enzymol.* 155:51-93 (1987)). Approximately three hours are required to prepare the agarose gel, electrophorese the sample, electroelute fragments larger than 500 bp, perform phenol-chloroform extractions, and precipitate the resulting material.

5 The results of 5 out of 9 independent trials size-fractionating CviJI^{**}-fragmented lambda DNA by agarose gel electrophoresis are shown in Figures 9A-E. Figures 9A-D illustrate the following. In Figure 9A: Lane M, 1 kb DNA ladder; lane λ , untreated λ DNA (0.25 μ g); lane 1, unfractionated (UF) CviJI^{**} partially-digested λ DNA (1.0 μ g); lane 2, column-fractionated (CF) CviJI^{**} partially-digested λ DNA (1.0 μ g); lane 3, gel-fractionated (GF) CviJI^{**} partially-digested λ DNA (1.0 μ g); and in Figures 9B-E are additional trials of the same treatments as in the lanes of Figure 9A which have the same label.

15 Small DNA fragments may also be removed by passing the sample through a short column of Sephacryl S-500. Approximately 15 min. are needed to prepare the column and 5 min. to fractionate the DNA by this method.

20 The results of three out of nine trials using a Sephacryl S-500 column are shown in Figures 9A-C. The efficiency of eliminating small DNA fragments (<500 bp) by spin column chromatography appears high, and the reproducibility was excellent. This result is in contrast to the agarose gel electrophoresis and electroelution data presented in Figures 9A-E wherein nine replicate trials of this method yielded nine differently sized products, regardless of the source of the agarose. Both methods yielded 30-40% recoveries as measured by UV spectrophotometry. To quantitate the relative efficiencies of the two fractionation methods, the lambda DNA size fractionated in Figure 9A lanes 2 and 3, and Figure 9B lane 3 were analyzed for cloning efficiency and insert size, as described below.

- 59 -

Example 15
Cloning Efficiencies of Gel Elution and
Chromatography Fractionation Methods

The efficacy of size selection was quantified by two criteria: 1) by
5 comparing the relative cloning efficiency of *Cvi*II** partially-digested lambda
DNA fragments fractionated either by agarose gel electrophoresis and
electroelution or micro-column chromatography, and 2) determining the size
distribution of the resulting cloned inserts. To reduce potential variables, large
quantities of the cloning vector and ligation cocktail were prepared, ligation
10 reactions and transformation of competent *E. coli* were performed on the same
day, numerous redundant controls were performed, and all cloning experiments
were repeated twice. Ligation reactions were carried out overnight at 12°C in 20
μl mixtures using the following conditions: 25 mM Tris-HCl (pH 7.8), 10 mM
MgCl₂, 1 mM DTT, 1 mM ATP, DNA, and 2000 units of T4 DNA ligase. For
15 unfractionated samples, 10 ng of fragments and 100 ng of *Hinc*II-restricted,
dephosphorylated pUC19 were combined under the above conditions. For
Sephacryl S-500 fractionated samples, 50 ng of size-selected fragments were
ligated with 100 ng of *Hinc*II-restricted, dephosphorylated pUC19. This increase
in fractionated DNA was determined empirically to compensate for the lower
20 concentration of "ends" resulting from the fractionation procedure and/or the
lowered efficiency of cloning larger fragments. Ligation reaction products were
added to competent *E. coli* DH5αF' (φ80*dlacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR*
gyrA96 recA1 relA1 endA1 thi-1 hsdR17(r_K⁻,m_K⁺) *supE44* λ-) in a
transformation mixture as specified by the manufacturer (Life Technologies,
25 Bethesda, Maryland) and aliquots of the transformation mixture were plated on
T agar (Messing, *Methods in Enzymol.* 101:20-78 (1983)) containing 20 μg/ml
ampicillin, 25 μl of a 2% solution of isopropylthiogalactoside (IPTG) and 25 μl
of a 2% solution of 5-dibromo-4-chloro-3-indolylgalactoside (X-GAL). The

- 60 -

cloning efficiencies reported are the average of triplicate platings of each ligation reaction. The concentration of the fractionated material was checked spectrophotometrically so that 50 ng was added to all ligation reactions. This material was ligated to *Hinc*II-digested and dephosphorylated pUC19. This

5 cloning vector was chosen because it permits a simple blue to white visual assay to indicate whether a DNA fragment was cloned (white) or not (blue) (Messing, *Methods in Enzymol.* 101:20-78 (1983)).

A summary of the cloning efficiencies calculated from two independent trials is given in Table 3.

- 61 -

TABLE 3

Cloning Efficiencies of CviJI^{**} Partially Digested Lambda DNA
 Fractionated by Microcolumn Chromatography Versus Agarose Gel
 Electroelution.

5		Trial I		Trial II	
		<u>Colony Phenotype</u>			
	<u>DNA/treatment</u>	<u>Blue</u>	<u>White</u>	<u>Blue</u>	<u>White</u>
	Supercoiled pUC19	55000	< 10	50000	< 10
	pUC19/HincII/CIAP	210	< 1	320	1
10	pUC19/HincII/CIAP/ T4 DNA ligase	150	4	210	7
	λ/CviJI ^{**} partial/CF + pUC19	140	240	210	240
	λ/CviJI ^{**} partial/GFE1 + pUC19	98	49	200	18
15	λ/CviJI ^{**} partial/GFE2 + pUC19	82	54	95	74

20 Cloning efficiencies reflect the number of ampicillin-resistant colonies/ng pUC19 DNA. CIAP represents treatment with calf intestinal alkaline phosphatase used to dephosphorylate *HincII*-digested pUC19 to minimize self-ligation. CF refers to DNA that was fractionated on Sephacryl S-500 columns as described above. GFE1 and GFE2 refer to two runs wherein DNA was fractionated by agarose gel electrophoresis and electroeluted. λ refers to bacteriophage λ DNA.

25 These trials represent repeated experiments in which λ DNA fragments generated by CviJI^{**} partial digestion were ligated to *HincII*-linearized,

- 62 -

dephosphorylated pUC19 and transformed into DH5 α F' competent cells described above. The first three rows in Table 2 show controls performed to establish a baseline to better evaluate the various treatments. Supercoiled pUC19 transforms *E. coli* 10 times more efficiently than the *HincII*-digested plasmid and 150-260 times more efficiently than the *HincII*-digested and dephosphorylated plasmid. The number of blue and white colonies which resulted from transforming *HincII*-cut and dephosphorylated pUC19 was determined both before and after treatment with T4 DNA ligase in order to differentiate these background events from cloning inserts. The background of blue colonies (which represent the uncut and/or non-dephosphorylated population of molecules) averaged 0.4%, compared to supercoiled plasmid. The background of white colonies (which presumably results from contaminating nucleases in the enzyme treatments or genomic DNA in the plasmid preparations) after *HincII*-digestion, dephosphorylation, and ligation of pUC19 averaged 0.014% as compared to the supercoiled plasmid.

The number of white colonies obtained when micro-column fractionated DNA was cloned into pUC19 was 240/ng vector in both trials. The efficiency of cloning gel fractionated and electroeluted DNA ranged from 18-74 white colonies/ng vector. The data show that column fractionated DNA results in three to thirteen times the number of white colonies, and presumably recombinant inserts, as gel fractionated and electroeluted DNA. The size distribution of the inserts present in these white colonies is depicted in Figures 10A-C. In Figure 10A, a *Cvi*II** partial digest of 2 μ g of λ DNA was size fractionated on a 4 mm by 13 mm column of Sephacryl S-500 at 2,000 x g for 5 minutes. The void volume containing partially digested DNA was directly ligated to linear, dephosphorylated pUC19 and 43 resulting clones were analyzed for insert size. The DNA for this experiment is the same as that shown in Figure 9A, lane 2. In Figure 10B, a *Cvi*II** partial digest of 5 μ g of λ DNA was size fractionated by agarose gel electroelution. The eluted DNA was phenol-extracted and ligated to linear, dephosphorylated pUC19, and the resulting 40 clones were

- 63 -

analyzed for insert size. The DNA for this experiment is the same as that shown in Figure 9A, lane 3. In Figure 10C, the procedure is the same as in Figure 9B, except the DNA for this experiment came from Figure 9B, lane 3.

5 A total of 43 random clones obtained from micro-column chromatography fractionation were analyzed for insert size (as shown in Figure 10A). Most of these inserts were larger than 500 bp (37/43 or 86%), 11.6% (5/43) were smaller than 500 bp, and one clone (2.3%) was smaller than 250 bp. The average insert size was 1630 bp. These results are in contrast to those obtained by agarose gel fractionation (as shown in Figures 10B and 10C). In the first trial (Figure 10B) most of the inserts were smaller than 500 bp (26/37 or 70.3%) and only 29.7% (11/37) were larger than 500 bp in size. In the second trial (Figure 10C) all of the inserts (40 total) were smaller than 500 bp. Thus, the use of agarose gel electroelution for the size fractionation of DNA results in unexpectedly variable and low cloning efficiencies.

15

Example 16

Cloning Sonicated and *Cvi*JI^{**}-Digested Lambda DNA

To compare the cloning efficiencies of sonicated and *Cvi*JI^{**}-digested nucleic acid, λ DNA was fragmented by each of these methods and ligated to pUC19 which was linearized with *Hinc*II and dephosphorylated to minimize self-ligation.

20

DNA fragmented by *Cvi*JI^{**} digestion and sonication was cloned both before and after Sephacryl S-500 size fractionation. Sonicated lambda DNA was subjected to an end repair treatment prior to ligation. Ligations were performed as described in Example 11. One-tenth of the ligation reaction (2 μ l) was utilized in the transformation procedure, and the fraction of nonrecombinant (blue) versus recombinant (white) colonies was used to calculate the efficiency of this process.

25

- 64 -

The efficacy of the methods was quantified by comparing the cloning efficiency of lambda DNA fragments generated either by sonication or *Cvi*II** partial digestion. To reduce potential cloning differences based on size preference, the size distribution of the DNA generated by these two methods was closely matched. Other experimental details were designed to reduce potential variables, as described above. Certain variables were unavoidable, however. For example, the sonicated DNA fragments required an enzymatic step to repair the ragged ends as described in Example 1 prior to ligation, whereas the *Cvi*II** digests were heat-denatured and directly ligated to *Hinc*II digested pUC19.

10 A summary of the cloning efficiencies calculated from two independent trials is given in Table 4, section A (unfractionated samples), and Section B (fractionated samples).

TABLE 4
Cloning Efficiencies of *Cvi*II** Partially Digested λ DNA
Versus Sonicated λ DNA

A. Unfractionated Samples

DNA/treatment	Trial I		Trial II	
	Colony Phenotype		Colony Phenotype	
	Blue	White	Blue	White
Supercoiled pUC19	30000	< 10	16000	< 10
pUC19/ <i>Hinc</i> II/CIAP	150	< 1	31	1
pUC19/ <i>Hinc</i> II/CIAP/ T4 DNA ligase	100	< 1	15	1
λ /AluI + pUC19	200	400	73	250
λ / <i>Cvi</i> II** Partial + pUC19	100	160	97	340
λ /Sonicated + pUC19	-	-	11	29
λ /Sonicated/ER 1 + pUC19	17	10	10	44
λ /Sonicated/ER 2 + pUC19	-	-	40	100

6 6

TABLE 4 (cont'd)

B. Fractionated Samples

DNA/treatment	Trial I		Trial II	
	Colony Phenotype		Colony Phenotype	
	Blue	White	Blue	White
Supercoiled pUC19	35000	<10	12000	<10
pUC19/ <i>HincII</i> /CIAP	30	<1	180	<1
pUC19/ <i>HincII</i> /CIAP/ T4 DNA ligase	60	<1	10	<1
λ / <i>Acl</i> I + pUC19	28	23	33	48
λ / <i>Cvi</i> II** Partial + pUC19	31	90	36	68
λ /Sonicated + pUC19	20	6	99	19
λ /Sonicated/ER 1 + pUC19	27	32	40	19
λ /Sonicated/ER 2 + pUC19	-	-	25	63

- 67 -

Cloning efficiencies represent the number of ampicillin-resistant colonies/ng pUC19 DNA. CIAP indicates treatment with calf intestinal alkaline phosphatase. ER 1 and ER 2 are end repair methods described in Example 13. λ refers to bacteriophage lambda.

5 The indicated trials represent repeated experiments in which two identical sets of lambda DNA fragments generated by *AluI* complete digestion, *CviJI*^{**} partial digestion, or sonication were each ligated to *HincII*-linearized, dephosphorylated pUC19 and transformed into DH5 α F' competent cells. The cloning efficiencies reported are the average of triplicate platings of each ligation
10 reaction. In case the Sephacryl S-500 size fractionation step introduced inhibitors of ligation or transformation or resulted in differences attributable to the size of the material, the sonicated and *CviJI*^{**}-digested samples were ligated with pUC19 both prior to (A) and after (B) the fractionation steps. The first three rows in Table 4, sections A and B, are controls performed to establish a baseline to better
15 evaluate the various treatments. These data show that supercoiled pUC19 transforms *E. coli* 200-1000 times more efficiently than the *HincII*-restricted and dephosphorylated plasmid. Without this dephosphorylation step, the cloning efficiency is 10% that of the supercoiled molecule (data not presented). The background of blue colonies averaged 0.5% in these experiments, compared to
20 supercoiled plasmid, while the background of white colonies averaged 0.005%.

A comparison of the data from unfractionated versus fractionated samples in Table 4, sections A and B, reveals a general decline in the number of white and blue colonies obtained after sizing. This decrease is primarily due to the fact that cloning efficiencies are dependent upon the size of the fragment,
25 favoring smaller fragments and thus giving higher efficiencies for the unfractionated material. This is illustrated by comparing the efficiency of cloning unfractionated and fractionated λ DNA which was completely restricted with *AluI*. This four base recognition endonuclease produces blunt ends and cuts λ DNA (48,502 bp) at 143 sites. Only 25 of the resulting 144 fragments (17%) are larger

- 68 -

than 500 bp. The number of white colonies obtained when unfractionated λ DNA, completely restricted with *AluI*, was cloned into pUC19 ranged from 250-400/ng vector, versus 23-48/ng vector for the fractionated material. This ten fold decrease was only noticed for the λ *AluI* digests, and probably reflects the large portion of small molecular weight fragments (approximately 75%) which is excluded from the fractionated ligation reactions.

The number of white colonies obtained when unfractionated *CviJI*^{**} treated λ DNA was cloned into pUC19 ranged from 160-340/ng vector, versus 68-90 white colonies/ng vector if the same material was fractionated. Unfractionated λ DNA, completely digested with *AluI*, results in cloning efficiencies very similar to unfractionated *CviJI*^{**} treated DNA. Sonicated λ DNA is a poor substrate for ligation, compared to *CviJI*^{**} treatment, as indicated by the roughly ten-fold reduced cloning efficiencies.

Enzymatic repair of the ragged ends produced by sonication results in an increased cloning efficiency. Using conditions described in Example 13 for the first end repair treatment (ER 1), 10-44 (fractionated) and 19-32 (unfractionated) white colonies/ng vector were observed. However, ER 1 conditions may not be optimal, as an alternate end repair reaction (ER 2) (as described in Example 13) resulted in greater numbers of white colonies (63 and 100/ng vector for fractionated and unfractionated DNA, respectively). In this reaction, a ten-fold excess of reagents and enzymes were utilized to repair the sonicated DNA, which apparently improved the efficiency of cloning such molecules by two to three fold. The data collected from multiple cloning trials in Table 3, sections A and B, show that *CviJI*^{**} partial digestion results in three to sixteen times the number of white colonies than sonicated ER 1-treated DNA. Even with an optimal end repair reaction for the sonicated fragments, DNA treated with *CviJI*^{**} yielded three times more white colonies.

- 69 -

Example 17

Analysis of *Cvi*JI** Fragmentation for Shotgun Cloning and Sequencing

The ability of *Cvi*JI** partial digestion to create uniformly representative clone libraries for DNA sequencing was tested on pUC19 DNA. pUC19 DNA was digested under *Cvi*JI** conditions and size fractionated as described above. The fractionated DNA was cloned into the *Eco*RV site of M13SPSI, a lacZ minus vector constructed by adding an *Eco*RV restriction site to wild type M13 at position 5605. M13SPSI lacks a genetic cloning selection trait, therefore after ligation of the pUC19 fragments into the vector the sample was restricted with *Eco*RV to reduce the background of nonrecombinant plaques. Bacteriophage M13 plaques were picked at random and grown for 5-7 hours in 2 ml of 2XTY broth containing 20 μ l of a DH5 α F' overnight culture. After centrifugation to remove the cells, single-stranded phage DNA was purified using Sephaglass™ as specified by the manufacturer (Pharmacia LKB, Piscataway New Jersey). The single-stranded DNA was sequenced by the dideoxy chain termination method using a radiolabeled M13-specific primer and *Bst* DNA polymerase (Mead *et al.*, *Biotechniques* 11:76-87 (1991)). The first 100 bases of 76 randomly chosen clones were sequenced to determine which *Cvi*JI recognition site was utilized, the orientation of each insert and how effectively the cloned fragments covered the entire molecule, as shown in Figure 11. The positions of the 45 normal *Cvi*JI sites (PuGCPy) in pUC19 are indicated beneath the line labeled "NORMAL" in the Figure 11. Similarly, the 160 *Cvi*JI* sites (GC) are indicated beneath the line labeled "RELAXED" in Figure 11. The marks above these lines indicate the *Cvi*JI** pUC19 sites which were found in the set of 76 sequenced random clones. The frequency of cloning a particular site is indicated by the height of the line, and the left or right orientation of each clone is also indicated at the top of each mark. There are a total of 205 *Cvi*JI and *Cvi*JI* sites in pUC19.

- 70 -

The data presented in Figure 11 demonstrate that, under *Cvi*II^{**} partial conditions, normal *Cvi*II sites are preferentially restricted over relaxed (*Cvi*II^{*}) sites. Of the 76 clones that were analyzed, only 13%, or 1 in 7, had sequence junctions corresponding to a relaxed *Cvi*II^{*} site. Thirty-five of the
5 forty-five possible normal restriction sites were cloned, as compared to eight of the possible one hundred sixty relaxed sites. If the enzyme had exhibited no preference for normal or relaxed sites under the *Cvi*II^{**} partial conditions utilized here, then 78% of the sequence junctions analyzed should have been generated by cleavage at a relaxed *Cvi*II^{*} site. It may be noted that the relaxed *Cvi*II^{*}
10 restriction sites that were found appear to be clustered in two regions of the plasmid that are deficient in normal *Cvi*II sites. In addition, the combined distribution of the normal and relaxed sites which were restricted to generate the 76 clones appears to be quasi-random. That is, the longest gap between cloned restriction sites was no greater than 250 bp and no one particular site is over-
15 utilized.

A detailed analysis of the distribution of *Cvi*II^{**} sequence junctions found from cloning pUC19 is presented in Table 5.

TABLE 5
Distribution of Cloned CviJI** Partially-Digested pUC19 Sites.

Classification Group	Recognition Sequence	NGCN		Cloned CviJI** Distribution (%)	Pu/Py Structure
		Site Distribution in pUC19 (%)			
Normal (N)	A C	AGCC 9	(4.4)	13(17.1)	PuPuPyPy
	GC	GGCC 11	(5.4)	16(21.1)	
	G T	GGCT 10	(4.9)	12(15.8)	
		AGCT 15	(7.3)	25(32.9)	
		45	(22.0)	66(86.9)	
Relaxed (R ₁)	C C	CGCC 11	(5.4)	0	PyPuPyPy
	GC	TGCC 12	(5.9)	2 (2.6)	
	T T	TGCT 10	(4.9)	1 (1.3)	
		CGCT 22	(10.7)	2 (2.6)	
		55	(26.9)	5 (6.5)	

TABLE 5 (cont'd)

Classification Group	Recognition Sequence	NGCN		Cloned CviJI** Distribution (%)	Pu/Py Structure
		Site Distribution in pUC19 (%)			
Relaxed (R ₂)	A A	AGCA 16	(7.3)	1 (1.3)	PuPuPyPu
	GC	GGCA 8	(3.9)	0	
	G G	AGCG 11	(5.4)	0	
		GGCG 22	(10.7)	4 (5.2)	
		57	(27.8)	5 (6.6)	
Relaxed (R ₃)	C A	CGCA 10	(4.9)	0	PyPuPyPu
	GC	TGCA 13	(6.3)	0	
	T G	CGCG 10	(4.9)	0	
		TGCG 15	(7.3)	0	
		48	(23.4)	0	

- 73 -

The GC sites in pUC19 may be divided into four classes based on their flanking Pu/Py structure. The fraction of GC sites observed in pUC19 which belong to each classification is roughly equal (22.0-27.8%). A striking difference was found between the observed distribution in pUC19 of normal and relaxed (R1, R2, R3) CviJI recognition sites and the distribution revealed by shotgun cloning and sequence analysis of CviJI^{**}-treated DNA. While most of the sites cleaved by this treatment were found to be PuGCPy (about 87%), or "normal" restriction sites, a significant fraction of the cleavage occurred at PyGCPy (about 6.5%) and PuGCPu (about 6.6%) sites, considering the short incubation times and limiting enzyme concentrations. The latter two categories of sites, and presumably the PyGCPu sites as well, are completely restricted under "relaxed" conditions, provided an excess of enzyme is present and sufficient time is allowed (see Figure 7, and Xia *et al.*, *Nucleic Acids Res.* 15:6075-6090 (1987)).

Digestion using CviJI^{**} treatment results in a relatively even distribution of breakage points across the length of the molecule (as shown in Figure 11). As described above, Figure 11 depicts a linear map of pUC19 showing the relative position of the lacZ' gene (α peptide of β -galactosidase gene) and ampicillin resistance gene (Amp). The marks extending beneath the top line (labeled "NORMAL") show the relative position of the 45 normal CviJI sites (PuGCPy) present in pUC19. The marks above the line are the cleavage sites found from sequencing the CviJI^{**} partial library. The height of the line indicates the number of clones obtained from cleavage at that site, and the orientation of the flag designates the right or left orientation of the respective clone. The marks extending beneath the second line (labeled "RELAXED") show the relative positions of the 160 CviJI^{*} sites (GC) present in pUC19. Those marks above the line were found from sequencing the CviJI^{**} partial library. The bottom portion of Figure 11 shows the relative position and orientation of the first 20 clones sequenced, assuming a 350 bp read per clone. CviJI^{**} cleavage at relaxed sites appears to be important in "filling gaps" left by normal restriction.

- 74 -

The primary goal of this effort was to determine the efficacy of these methods for rapid shotgun cloning and sequencing. For these purposes, only 100 bases of sequence data were acquired per clone. However, if 350 bases of sequence had been determined from each clone, then the entire sequence of pUC19 would have been assembled from the overlap of the first 20 clones (Figure 11). In this sequencing simulation 75% of pUC19 would have been sequenced at least 2 times from the first 20 clones. The highest degree of overfold sequencing would have been 6, and only involved 2.2% of the DNA. Figure 11 also shows that most of the 1x sequencing coverage occurred in a region of the plasmid with a very low density of normal and relaxed *Cvi*JI restriction sites. Most of the single coverage occurs in a 240 bp region of the plasmid between 1490 bp and 1730 bp where there are only 4 *Cvi*JI relaxed sites. It should also be noted that by the 27th randomly picked clone most of this region would have been covered a second time.

Shotgun sequencing strategies are efficient for accumulating the first 80-95% of the sequence data. However, the random nature of the method means that the rate at which new sequence is accumulated decreases as more clones are analyzed. In Figure 12 the total amount of unique pUC19 sequence accumulated was plotted as a function of the number of clones sequenced. The points represent a plot of the total amount of determined pUC19 sequence versus the total number of clones sequenced. The horizontal dashed line demarcates the 2686 bp length of pUC19. The smooth curve represents a continuous plot of the discrete function $S(N) = NLe^{-c\sigma}[(e^{c\sigma}-1)/c + (1-\sigma)]$. The theoretical accumulation curve expected for a process in which sequence information is acquired in a totally random fashion is also shown. The smooth curve is a continuous plot of the discrete function $S(N)$ where

$$S(N) = NLe^{-c\sigma}[(e^{c\sigma}-1)/c + (1-\sigma)].$$

This equation is based upon the results developed by Lander *et al.*, *Genomics* 2:231-239 (1988) for the progress of contig generation in genetic mapping. In the

- 75 -

equation: N is the number of clones sequenced, L is the length of clone insert in bp, c is the redundancy of coverage or LN/G (where G is length of fragment being sequenced in bp), and $\sigma = 1 - \Theta$, where Θ is the fraction of length that two clones must share. The curve in Figure 12 was calculated with $G = 2686$ bp, $L = 350$ bp, and $\sigma = 1$. The plotted points lie close to the theoretical curve, and it thus appears that the sequence of pUC19 was accumulated in an apparent random fashion utilizing *Cvi*JI^{***} fragmentation and column fractionation.

Example 18

Shotgun Cloning Utilizing 200 ng of Lambda DNA

Generally, 2-5 μ g of DNA are needed for the sonication and agarose gel fractionation method of shotgun cloning in order to provide the several hundred colonies or plaques required for sequence analysis (Bankier *et al. Methods in Enzymol.* 155:51-93 (1987)). A ten-fold reduction in the amount of substrate required greatly simplifies the construction of such libraries, especially from large genomes, (Davidson, *J. DNA Sequencing and Mapping* 1:389-394 (1991)). The efficiency of constructing a large shotgun library from nanogram amounts of substrate was tested utilizing 200 ng of *Cvi*JI^{***}-digested lambda DNA. This material was column-fractionated as described previously. In this case, 1/2 of the column eluant (15 μ l containing 50 ng of DNA) was ligated to 100 ng of *Hinc*II-digested and dephosphorylated pUC19 as described in Example 15. The cloning efficiencies of the control DNAs were similar to those reported in Tables 2 and 3. The 50 ng cloning experiment yielded 230 white colonies per ligation reaction in one trial, and 410 white colonies per ligation reaction in a second trial. Thus, it should be possible to routinely construct useful quasi-random shotgun libraries from as little as 0.2 - 0.5 μ g of starting material.

- 76 -

Example 19

Epitope Mapping

*Cvi*JI* recognizes the sequence GC (except for PyGCPu) in the target DNA. Under partial restriction conditions the length of fragment may be controlled by incubation time. Epitope mapping using *Cvi*JI** partial digests involves generating DNA fragments of 100-300 bp from a cDNA coding for the protein of interest, by methods described in Example 13, inserting them into an M13 expression vector, plating out on solid media, lifting plaques onto a membrane, screening for binding to the ligand of interest, and picking the positive plaques for isolation of the DNA, which is then sequenced to identify the epitope. Thus, the same epitope may be expressed as a small fragment or a larger fragment. This approach allows one to determine the smallest fragment containing the epitope of interest using functional assays such as binding to an antibody or other ligand, or using a direct assay for activity. For insertion into an M13 vector, linkers may be added to the fragments or the insert may be dephosphorylated to ensure that each fragment is cloned alone without ligation of multiple inserts.

The expression vectors recommended for subcloning of the *Cvi*JI fragments are Lambda Zap[™] (Stratagene, LaJolla, California) or bacteriophage M13-epitope display vectors. An advantage of using an M13-based vector is that the peptide or protein of interest may be displayed along with the M13 coat protein and does not require host cell lysis in order to analyze the protein of interest. The lambda-based vectors yield plaques and hence the protein can be directly bound to a membrane filter.

- 77 -

Example 20**CGase I**

CGase I as used herein, refers to a restriction endonuclease reagent which cleaves DNA at the dinucleotide CG. CGase I activity is based on the combined
5 star activities of the restriction endonucleases Hpa II and Taq I. Under normal reaction conditions (10 mM Bis Tris Propane-HCl pH 7.0, 10 mM MgCl₂, 1 mM DTT; 1 unit of enzyme/ μ g DNA, 37°C for 1 hr), Hpa II recognizes CCGG and cleaves after the first C to leave a 2-base 5' overhang. Under normal reaction
10 conditions (100 mM NaCl, 10 mM Tris-HCl pH 8.4, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 1 unit of enzyme/ μ g DNA, 65°C for 1 hr) the restriction endonuclease Taq I recognizes TCGA and cleaves after the T to leave a 2-base 5' overhang.

Reaction conditions have been described for Taq I* activity which decrease the cleavage specificity of Taq I (10 mM Tris-HCl pH 9.0, 5 mM MgCl₂, 6 mM
15 2-mercaptoethanol, 20% DMSO; 2000 units of enzyme/ μ g DNA, 65°C for 1 hr) (Barany, Gene, 65:149-165 (1988)). These reaction conditions allow Taq I* to cleave DNA at the following sequences:

20 Taq I* TCGA
 CCGA (TCGG)
 ACGA (TCGT)
 TCTA (TAGA)
 TCAA (TTGA)
 GCGA (TCGC)

25 We are unaware of any literature descriptions of Hpa II* conditions. However, the following conditions were established to promote Hpa II* activity which are also compatible with Taq I* activity: 5 mM KCl, 10 mM Tris-HCl pH 8.5, 10 mM MgCl₂, 1 mM DTT, 15% DMSO, 100 μ g/ml BSA (CGase buffer); 50 units of enzyme/ μ g DNA 50°C for 1 hr. The Hpa II* recognition sites were

- 78 -

determined by cloning and sequencing Hpa II* restricted fragments. The characterized Hpa II* recognition sequences are as follows:

5 Hpa II* CCGG
CCGC (GCGG)
CCGA (TCGG)
ACGG (CCGT)

10 Taq I (400 units/ μ g DNA) and Hpa II (50 units/ μ g DNA) were then combined (CGase I) in CGase I buffer and the following recognition sites were identified by cloning and sequencing restricted pUC19 fragments.

15 CGase I GCGC
TCGA
CCGG
GCGT
ACGA
ACGG (CCGT)
GCGG (CCGC)
CCGA (TCGG)

20 CGase I restriction of natural DNA, (i.e. pUC19, lambda), results in fragments ranging from 20-200 bp in length (average 20-60 bp). Heat denaturation of these fragments generates numerous oligonucleotides of variable length but precise specificity for the cognate template as was the case with CviI I* digestion. CGase I restriction of the small plasmid pUC19 (2689 bp) theoretically yields 174 restriction fragments, or 384 oligonucleotides after a heat denaturation step.

25 The "two-cutter" activity of CviI I* and CGase I represent a unique class of restriction endonuclease activity in that no other known restriction endonucleases will generate this size range of oligonucleotides. The ability to generate numerous oligonucleotides with perfect sequence specificity from any DNA, without regard to sequence composition, genetic origin, or prior sequence
30 knowledge is one of the properties that CGase I shares with CviI I*. In addition, the generation of numerous oligonucleotides by CviI I or CGase I results in a

- 79 -

form of probe or primer amplification not practical using conventional means of organic synthesis.

Based on ability to recognize a dinucleotide sequence, the present invention contemplates the interchangeability of CGase I with CviJ I* in all of the applications described herein.

Example 21

Purification of CviJ I Restriction Endonuclease from IL-3A-Infected *Chlorella* Cells

CviJ I was prepared by a modification of the method described by Xia *et al.*, *Nucl. Acids Res.* 15:6025-6090 (1987). *Chlorella* NC64A cells (ATCC Accession No. 75399 deposited on January 21, 1993, American Type Culture Collection, Rockville, Maryland) were infected with the virus IL-3A (ATCC Accession No. 75354 deposited November 6, 1992, American Type Culture Collection, Rockville, Maryland) according to Van Etten *et al.*, *Virology* 126:117-125 (1983). Five grams of IL-3A infected *Chlorella* NC64A cells were suspended in a glass homogenization flask with 15 g of 0.3 mm glass beads in buffer A (10 mM Tris-HCl pH 7.9, 10 mM 2-mercaptoethanol, 50 μ g/ml phenylmethylsulfonyl fluoride (PMSF), 20 μ g/ml benzamidine, 2 μ g/ml o-phenanthroline). Cell lysis was carried out at 4000 rpm for 90 sec in a Braun MSK mechanical homogenizer (Allentown, PA) with cooling from a CO₂ tank. After lysis 2 M NaCl was added to a final concentration of 200 mM, after which 10% polyethyleneimine (PEI) (Life Technologies, Bethesda, MD) (pH 7.5) was added to a final concentration of 0.3%. The mixture was then stirred for 2 hrs. at 4°C then centrifuged for 1 hr. at 50,000 g. Ammonium sulfate was added to the supernatant to 70% saturation and stirred overnight. A protein pellet was recovered by centrifugation for 1 hr. at 50,000 g. The resulting pellet was dissolved in 20 ml of buffer B (20 mM Tris-acetate pH 7.5, 0.5 mM EDTA, 10

- 80 -

mM 2-mercaptoethanol, 10% glycerol, 30 mM KCl, 50 ug/ml PMSF, 20 μ g/ml benzamidine [Sigma, St. Louis, Missouri], 2 μ g/ml o-phenanthroline [Sigma]) and dialysed against 500 ml of buffer B with 3 changes. The dialysed solution was then applied to 1 x 6 cm Heparin-Sepharose (Pharmacia LKB, Piscataway, New Jersey) column. After a 50 ml wash with buffer B, a 100 ml gradient of 0 to 0.7 M KCl in buffer B was run. Fractions having CviI I activity as measured by digestion of pUC19 DNA and agarose gel electrophoresis, were pooled, diluted in 5 volumes of buffer C (10 mM K/PO₄ pH 7.4, 0.5 mM EDTA, 10 mM 2-mercaptoethanol, 75 mM NaCl, 0.05% Triton X-100, 10% glycerol, 50 μ g/ml PMSF, 20 μ g/ml benzamidine, 2 μ g/ml o-phenanthroline) and applied to a 1 x 7 cm Phosphocellulose P11 (Whatman) column equilibrated in buffer C. After washing with 30 ml of buffer C, CviI I was eluted by a 100 ml gradient of 0 to 0.7 M NaCl in buffer C. At this step CviI I activity separated from non-specific nucleases. CviI I containing fractions were pooled and diluted in 4 volumes of buffer C and applied to a 1 x 4 cm hydroxyapatite HTP column (BioRad, Hercules, CA). After washing with 30 ml of buffer C, CviI I was eluted by a 0 to 0.7 M potassium phosphate (pH 7.4) gradient in buffer C. Active fractions containing CviI I activity and lacking non-specific nuclease activity were pooled and were dialysed overnight against storage buffer (50 mM potassium phosphate 200 mM KCl, 0.5 mM EDTA, 50% glycerol, 20 ug/ml PMSF were pooled) and stored at -20°C.

Although the present invention has been described in types of preferred embodiments, it is intended that the present invention encompass all modifications and variations which occur to those skilled in the art upon consideration of the disclosure herein, and in particular those embodiments which are within the broadest proper interpretation of the claims and their requirements.

- 81 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Molecular Biology Resources, Inc.
- (ii) TITLE OF INVENTION: Materials and Methods for
Restriction Endonuclease Applications
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 - (B) STREET: 6300 Sears Tower, 233 South Wacker Drive
 - (C) CITY: Chicago
 - (D) STATE: Illinois
 - (E) COUNTRY: United States of America
 - (F) ZIP: 60606-6402
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Clough, David W.
 - (B) REGISTRATION NUMBER: 36,107
 - (C) REFERENCE/DOCKET NUMBER: 28003/31967/PCT
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 312/474-6300
 - (B) TELEFAX: 312/474-0448
 - (C) TELEX: 25-3856

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAATTTCACA CAGGAAACAG CTATGTCTTT TCGCACGTTA GAAC

44

- 82 -

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGTCTTTTC GCACGTTAGA ACTATTCGCC GGTATAGCTG GTATTCACA TGGCCTCAGA	60
GGTATATCTA CACCAGTTGC ATTCGTAGAA ATTAATGAAG ACGCACAAAA ATTCTTGAAA	120
ACAAAGTTTT CAGATGCATC TGTATTCAAT GACGTTACGA AATTACCAA ATCGGACTTC	180
CCAGAAGACA TAGACATGAT TACTGCGGGA TTCCCGTGCA CTGGGTTTAG TATTGCAGGT	240
TCTAGAAGTG GATTTCGAACA CAAGGAATCC GGTCTCTTTG CTGATGTTGT GCGAATCACG	300
GAAGAGTATA AACCTAAAAT AGTGTTTTTG GAAAACTCCC ATATGTTGTC CCACACTTAC	360
AATCTCGATG TCGTCGTAAA AAAGATGGAT GAAATTGGTT ATTTCTGCAA GTGGGTAAC	420
TGTCGGGCAT CAATTATAGG AGCCCATCAT CAACGCCACC GGTGGTTTTG TCTCGCGATT	480
CGAAAAGATT ATGAACCAGA AGAAATAATT GTATCTGTGA ATGCTACAAA GTTCCACTGG	540
GAAAATAATG AACCACCGTG TCAAGTAGAC AATAAGAGTT ACGAGAATTC AACTCTTGTT	600
CGTCTGGCAG GATATTCCGT GGTCCCCGAC CAGATCAGAT ATGCTTTCAC CGGTCTATTT	660
ACAGGTGATT TTGAGTCATC GTGGAAAAC ACCTTGACAC CTGGGACAAT AATTGGCAGC	720
GAACACAAAA AAATGAAAGG AACTTACGAT AAAGTCATAA ACGGGTATTA TGAGAACGAT	780
GTGTATTATT CTTTTCAAG GAAAGAAGTT CATCGCGCTC CTCTAAATAT ATCCGTGAAA	840
CCACGTGATA TTCCGGAGAA ACATAACGGA AAAACACTCG TAGATCGCGA AATGATCAAG	900
AAATATTGGT GCACACCATG TGCTAGTTAT GGCCTGCTA CTGCTGGATG CAATGTTCTG	960
ACAGACCGTC AGTCACATGC ACTTCCTACA CAAGTCAGGT TTTCATATAG GGGTGTATGT	1020
GGACGACATT TGTCTGGTAT ATGGTGTGCA TGGTTGATGG GGTATGACCA AGAATATCTT	1080
GGTTATTGGG TTCAATATGA TAAAATATT TTGATACACT AAATGGATAT AAGAAGAAAA	1140
CGTTTTACAA TAGAAGGGGC TAAACGTATA ATACTCGAAA AAAAGAGACT TGAAGAGAAA	1200
AAAAGAATTG CGGAAGAGAA AAAAGAATT GCCTTATAG AAAACAACG AATTGCGGAA	1260
GAGAAAAAAA GAATTGCGGA AGAGAAAAAA CGATTGCGAC TTGAAGAGAA AAAACGAATT	1320
GCGGAAGAAA AAAAACGAAT CGCGGAAGAG AAAAAACGAA TCGTGGAACA GAAAAAAGA	1380
CTTGCACTTA TAGAAAAACA ACGAATTGCG GAAGAGAAAA TTGCGTCGGG GAGAAAAATT	1440

- 83 -

AGAAAGAGGA TCTCTACAAA TGCAACAAAA CATGAAAGAG AATTTGTCAA AGTTATAAAT	1500
TCAATGTTTCG TCGGACCCGC TACTTTTGTA TTCGTAGATA TAAAAGGTAA TAAATCCAGA	1560
GAAATCCACA ACGTTGTAAG ATTCAGACAA TTACAAGGCA GTAAAGCGAA ATCCCCGACC	1620
GCGTATGTTG ATAGAGAATA TAACAAACCT AAAGCGGATA TAGCAGCGGT AGACATAACC	1680
GGTAAAGATG TGGCATGGAT ATCCCATAAA GCATCTGAAG GATATCAACA ATATCTAAAA	1740
ATTTCTGGAA AGAACCTCAA GTTCACAGGA AAAGAATTAG AAGAAGTTCT ATCGTTCAAG	1800
AGAAAAGTAG TTAGTATGGC ACCGGTATCT AAAATATGGC CTGCTAATAA GACCGTATGG	1860
TCTCCTATCA AGTCAAATTT GATTAAAAAT CAAGCAATAT TCGGATTGA TTACGGTAAG	1920
AAACCAGGAA GGGACAATGT AGACATCATA GGTCAAGGAC GACCAATTAT AACAAAAAGA	1980
GGTTCCATAT TATATCTTAC ATTCACTGGT TTTAGCGCAT TAAATGGGCA CTTGGAGAAT	2040
TTTACTGGGA AACATGAACC CGTTTTCTAT GTAAGAACAG AACGGAGTAG TAGCGGGAGA	2100
AGTATAACAA CTGTCGTCAA TGGTGTCACT TATAAAAATT TAAGATTCTT TATACATCCA	2160
TACAACTTTG TTTCTTCAAA AACACAACGT ATTATGTAGG ACCATTTTCC CGAGAGACTT	2220
TGTTGACCGC GTACTAAAAA ATGGTCACGA TATTGTGCTA AAGATGCTCA TAGAAGCAGG	2280
TGCAACCTT GACATCGTCA GTGTTGAGTA TACACCATTA CATCTACATG TGGTGATATT	2340
TGTATAAACG GTAAATACCT ATATATACAA TACGTATCCC CCTAAAAGCG CTTAGATTTT	2400
TTAGTTGTAT ACTACTTTTG TATAAGACCT GTAAGTTACA AACTAAAGT TTCAGCTTTG	2460
CCTTCGAAAC AAGCAATTAC CGCATGAGAA TAATATCCAT TATGGATGTT TTCTGCTAAT	2520
AAAACGATAT TTCCTACAGA AGTTTTCTATG ATTAGTTCCG AAATATTGAG ATCATCGTCA	2580
CGTTTTTCTT TACCGTATTT TACTTTCGTG ATCGTCGCAC CAATAAAATC ATCTCGTGTG	2640
AGTTCATTCG GCAATTGTGC CGTGACACCA AATCTCTCAC AACACCTTG ATGTCCATCC	2700
ATTGCTAACA CTATCGGTAA TCCATGTGTG GTGTGTACGA CCACACCGTT ATAACATAA	2760
CACGTGTAGT TGTCGTCTAT ATCATATAAC TCGAGAGCGG TGTGAACCTC TTCAGATCTA	2820
TTATTAATCG GATCTGATCC ATAAGAAGAA TCTTCATATT TACAAATAAA ATCATCCGAT	2880
ATGTTCTGCA CACGAACAAC ATTCGTCAAA TTTCTGTGAT GACGAATCTC CATCTCTGAA	2940
TCATTAGAGA CTTGCGAGTA TATAACATTA TAATTGTTGA TATGATTATT ACGTTTCATA	3000
TCAACAAAAT ACATATAAAC ACCATACAAA TATTAAACA CGTTAGTATA TAATGGATAA	3060
CATTTGCAAT AGTATATTCA CTGCAGTAAA AAATGGCCAC GAAGCTTGTT TGAAGATGAT	3120
GCTCATTGAA AGAGGTAGCA ATATCAATGA TGTTTCCGAA TCAAAATATG GAAATACACC	3180
ACTACATATT GCAGCTCATC ATGGTAATGA TGTGTGTTT AAGATGCTTA TTGACGCAGG	3240
TGCAACCTT GATATCACAG ATATTTCTGG AGGAACACCA CTTTCATCGTG CGGTTTTGAA	3300

- 84 -

TGGCCATGAC ATATTGTACA GATGCTCGTA GAAGCAGGTG CAAACCTTAG TATCATAACT	3360
AATTTGGGAT GGATACCGTT ACATTACGCG GCTTTTAATG GTAATGATGC GATTTTGAGG	3420
ATGCTCATCG TTGTAAGTGA TAATGTTGAC GTTATCAATG ATCGCGGTTG GACGGCGTTA	3480
CATTACGCGG CTTTTAATGG TCATAGCATG TCGGTCAAGA CGCTTATTGA TCGGGGTGCA	3540
AATCTTGACA TCACAGATAT TTCGGGATGT ACACCACTTC ATCGTGCGGT TTATAATGAC	3600
CACGATGCAT GTGTGAAGAT ACTCGTAGAA GCAGGTGCAA CTCTTGACGT CATTGATGAT	3660
ACTGAGTGGG TGCCGTTACA TTACGCGGCT TTTAATGGTA ATGATGCGAT TTTGAGGATG	3720
CTCATTGAAG CAGGTGCAGA TATTGATATA TCTAATATAT GTGATTGGAC GGC GTTACAT	3780
TACGCGGCTC GAAATGGACA CGATGTGTGT ATAAAAACAC TCATCGAAGC AGGTGGTAAC	3840
ATCAACGCCG TCAACAAATC GGGGGATACA CCACTAGATA TTGCAGCATG TCATGACATT	3900
GCAGTATGTG TGATCGTGAT AGTCAATAAG ATCGTTTCGG AGCGGCCGTT GCGTCCGAGT	3960
GAGTTGTGTG TCATACCACC AACGTCTGCT GCATTAGGTG ATGTGTTGCG AACGACGATG	4020
CGGCTTCATG GCGGATCGGA AGCTGCAAAG ATCACAGCGC ATCTTCCTGT GGGTGCAAGG	4080
GATACTCTAC GAACTACTGC GTTGTGTTG AACCGAACAA TTTCCGAGAG ATCTCGTTGA	4140
TAGTGTATTA ATTGAATGCG TGTAAGTTA CGCTATTTTT TTCCAAAAG GGTTCGATG	4200
AAATACAACA CGATCTTTTG TAGATCGTTT ACCATTAGTT GTATTCGTGC AATAGAGACC	4260
ATACGTACCT CCAAATTCAT TTACTTTACC TACAGTATTA CCACTTCCTT TTTTTCCTAT	4320
AGTAGTATCT AAATTCAACC CTTTGAATC ATCGCCATTA ACAGACAGAG CGTATGAACC	4380
GTTTTGTGCC AATTTACCT TCAAAACGAT AGTAAACCAT TGACCTCTAG GAATTTTAAC	4440
CGATCTTATA AGTATCTGCT TACTTCCAAG TCCTTTTTCA AAAGCATACA ACGATCCTGT	4500
AAGGTTATCC CCAGAACCTG AAATTGTAAA GAACGACTGG AAATGAATAG GTTGCATTAG	4560
ATCTGTATAC ATATCACTTG GTTCGAAATG AAAATCGTAG TCCCAATTAG GTACGTTCCA	4620
CCAAGTTTAA TACGGGGTCT TTCCACCGAG ACCGGACATT TCAGCACGAG CCTTGTAAGA	4680
ATGATATGAT GTGGTTAAAT CTCTATCACC ATCGTTCCAC TTTCTCTGA ACCGAAGACC	4740
ATGCATCGTT ATACCTGGTG CAACCTGTAC TAAATTCTTT ATTTACAGGTG CGGCTCCGGG	4800
TGGATTAAC TCGAGATTCGT CAAATCTAAA ATATGATAAC GATGTTCCAA CAGTAGAACC	4860
ACTGGGTGCT ATGGCAGTTG CTGGAAGGGA AGGTAAACT TTAGGATATT TCAAATCACC	4920
AACACCTTGA GGGTTTACTT GAATACTTCT GGGAGATGTT GGTGGTTTCG TCGAAGGTGG	4980
TTTCGTTGAA GGTGGTTTCG TCGAAGGTGG TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG	5040
TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG	5100
TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG	5160

- 85 -

TTTCGTCGAA GGTGGTTTCG TCGAAGGTGG TTTCGTTGGC GGAAGTGGGG CATGACCATA 5220
 ATCCGTTAAA TTCCCGCATT CACCTAATGA TGTACTCCAT AAAGAACCGG GTGCGCATTG 5280
 CATTCTTATT GGTTCGTAG TATCAGATAT ACATACGAAA TAATGAGAAT CATTTCCTT 5340
 GCCAAATAAT TTACCAGATT TGCCTTTACA TGACATTATT TGTAATATAA TATTATTATA 5400
 ATTTTAAAAA AACTAACGTC TATTTAAAAT TATGTAATAC GTATTATATC AATGCATCAT 5460
 CTTAATCATT TCCTAACGTA TAAGCGTAGC GAATTC 5496

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1225 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(1..33, 55..1128)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CAA GAA TAT CTT GGT TAT TTG GTT CAA TAT GAT TAAAATATTT TGATACACTA 53
 Gln Glu Tyr Leu Gly Tyr Leu Val Gln Tyr Asp
 1 5 10

A ATG GAT ATA AGA AGA AAA CGT TTT ACA ATA GAA GGG GCT AAA CGT 99
 Met Asp Ile Arg Arg Lys Arg Phe Thr Ile Glu Gly Ala Lys Arg
 15 20 25

ATA ATA CTC GAA AAA AAG AGA CTT GAA GAG AAA AAA AGA ATT GCG GAA 147
 Ile Ile Leu Glu Lys Lys Arg Leu Glu Glu Lys Lys Arg Ile Ala Glu
 30 35 40

GAG AAA AAA AGA ATT GCA CTT ATA GAA AAA CAA CGA ATT GCG GAA GAG 195
 Glu Lys Lys Arg Ile Ala Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu
 45 50 55

AAA AAA AGA ATT GCG GAA GAG AAA AAA CGA TTC GCA CTT GAA GAG AAA 243
 Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Phe Ala Leu Glu Glu Lys
 60 65 70

AAA CGA ATT GCG GAA GAA AAA AAA CGA ATC GCG GAA GAG AAA AAA CGA 291
 Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg
 75 80 85 90

ATC GTG GAA GAG AAA AAA AGA CTT GCA CTT ATA GAA AAA CAA CGA ATT 339
 Ile Val Glu Glu Lys Lys Arg Leu Ala Leu Ile Glu Lys Gln Arg Ile
 95 100 105

GCG GAA GAG AAA ATT GCG TCG GGG AGA AAA ATT AGA AAG AGG ATC TCT 387
 Ala Glu Glu Lys Ile Ala Ser Gly Arg Lys Ile Arg Lys Arg Ile Ser
 110 115 120

- 86 -

ACA AAT GCA ACA AAA CAT GAA AGA GAA TTT GTC AAA GTT ATA AAT TCA Thr Asn Ala Thr Lys His Glu Arg Glu Phe Val Lys Val Ile Asn Ser 125 130 135	435
ATG TTC GTC GGA CCC GCT ACT TTT GTA TTC GTA GAT ATA AAA GGT AAT Met Phe Val Gly Pro Ala Thr Phe Val Phe Val Asp Ile Lys Gly Asn 140 145 150	483
AAA TCC AGA GAA ATC CAC AAC GTT GTA AGA TTC AGA CAA TTA CAA GGC Lys Ser Arg Glu Ile His Asn Val Val Arg Phe Arg Gln Leu Gln Gly 155 160 165 170	531
AGT AAA GCG AAA TCC CCG ACC GCG TAT GTT GAT AGA GAA TAT AAC AAA Ser Lys Ala Lys Ser Pro Thr Ala Tyr Val Asp Arg Glu Tyr Asn Lys 175 180 185	579
CCT AAA GCG GAT ATA GCA GCG GTA GAC ATA ACC GGT AAA GAT GTG GCA Pro Lys Ala Asp Ile Ala Ala Val Asp Ile Thr Gly Lys Asp Val Ala 190 195 200	627
TGG ATA TCC CAT AAA GCA TCT GAA GGA TAT CAA CAA TAT CTA AAA ATT Trp Ile Ser His Lys Ala Ser Glu Gly Tyr Gln Gln Tyr Leu Lys Ile 205 210 215	675
TCT GGA AAG AAC CTC AAG TTC ACA GGA AAA GAA TTA GAA GAA GTT CTA Ser Gly Lys Asn Leu Lys Phe Thr Gly Lys Glu Leu Glu Val Leu 220 225 230	723
TCG TTC AAG AGA AAA GTA GTT AGT ATG GCA CCG GTA TCT AAA ATA TGG Ser Phe Lys Arg Lys Val Val Ser Met Ala Pro Val Ser Lys Ile Trp 235 240 245 250	771
CCT GCT AAT AAG ACC GTA TGG TCT CCT ATC AAG TCA AAT TTG ATT AAA Pro Ala Asn Lys Thr Val Trp Ser Pro Ile Lys Ser Asn Leu Ile Lys 255 260 265	819
AAT CAA GCA ATA TTC GGA TTT GAT TAC GGT AAG AAA CCA GGA AGG GAC Asn Gln Ala Ile Phe Gly Phe Asp Tyr Gly Lys Lys Pro Gly Arg Asp 270 275 280	867
AAT GTA GAC ATC ATA GGT CAA GGA CGA CCA ATT ATA ACA AAA AGA GGT Asn Val Asp Ile Ile Gly Gln Gly Arg Pro Ile Ile Thr Lys Arg Gly 285 290 295	915
TCC ATA TTA TAT CTT ACA TTC ACT GGT TTT AGC GCA TTA AAT GGG CAC Ser Ile Leu Tyr Leu Thr Phe Thr Gly Phe Ser Ala Leu Asn Gly His 300 305 310	963
TTG GAG AAT TTT ACT GGG AAA CAT GAA CCC GTT TTC TAT GTA AGA ACA Leu Glu Asn Phe Thr Gly Lys His Glu Pro Val Phe Tyr Val Arg Thr 315 320 325 330	1011
GAA CGG AGT AGT AGC GGG AGA AGT ATA ACA ACT GTC GTC AAT GGT GTC Glu Arg Ser Ser Ser Gly Arg Ser Ile Thr Thr Val Val Asn Gly Val 335 340 345	1059
ACT TAT AAA AAT TTA AGA TTC TTT ATA CAT CCA TAC AAC TTT GTT TCT Thr Tyr Lys Asn Leu Arg Phe Phe Ile His Pro Tyr Asn Phe Val Ser 350 355 360	1107

- 87 -

TCA AAA ACA CAA CGT ATT ATG TAGGACCATT TTCCCGAGAG ACTTTGTTGA 1158
 Ser Lys Thr Gln Arg Ile Met
 365

CCGCGTACTA AAAAATGGTC ACGATAATTG TCTAAAGATG CTCATAGAAG CAGGTGCAAA 1218

CCTTGAC 1225

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 369 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gln Glu Tyr Leu Gly Tyr Leu Val Gln Tyr Asp Met Asp Ile Arg Arg
 1 5 10 15
 Lys Arg Phe Thr Ile Glu Gly Ala Lys Arg Ile Ile Leu Glu Lys Lys
 20 25 30
 Arg Leu Glu Glu Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala
 35 40 45
 Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu Lys Lys Arg Ile Ala Glu
 50 55 60
 Glu Lys Lys Arg Phe Ala Leu Glu Glu Lys Lys Arg Ile Ala Glu Glu
 65 70 75 80
 Lys Lys Arg Ile Ala Glu Glu Lys Lys Arg Ile Val Glu Glu Lys Lys
 85 90 95
 Arg Leu Ala Leu Ile Glu Lys Gln Arg Ile Ala Glu Glu Lys Ile Ala
 100 105 110
 Ser Gly Arg Lys Ile Arg Lys Arg Ile Ser Thr Asn Ala Thr Lys His
 115 120 125
 Glu Arg Glu Phe Val Lys Val Ile Asn Ser Met Phe Val Gly Pro Ala
 130 135 140
 Thr Phe Val Phe Val Asp Ile Lys Gly Asn Lys Ser Arg Glu Ile His
 145 150 155 160
 Asn Val Val Arg Phe Arg Gln Leu Gln Gly Ser Lys Ala Lys Ser Pro
 165 170 175
 Thr Ala Tyr Val Asp Arg Glu Tyr Asn Lys Pro Lys Ala Asp Ile Ala
 180 185 190
 Ala Val Asp Ile Thr Gly Lys Asp Val Ala Trp Ile Ser His Lys Ala
 195 200 205
 Ser Glu Gly Tyr Gln Gln Tyr Leu Lys Ile Ser Gly Lys Asn Leu Lys
 210 215 220

- 88 -

Phe Thr Gly Lys Glu Leu Glu Glu Val Leu Ser Phe Lys Arg Lys Val
 225 230 235 240
 Val Ser Met Ala Pro Val Ser Lys Ile Trp Pro Ala Asn Lys Thr Val
 245 250 255
 Trp Ser Pro Ile Lys Ser Asn Leu Ile Lys Asn Gln Ala Ile Phe Gly
 260 265 270
 Phe Asp Tyr Gly Lys Lys Pro Gly Arg Asp Asn Val Asp Ile Ile Gly
 275 280 285
 Gln Gly Arg Pro Ile Ile Thr Lys Arg Gly Ser Ile Leu Tyr Leu Thr
 290 295 300
 Phe Thr Gly Phe Ser Ala Leu Asn Gly His Leu Glu Asn Phe Thr Gly
 305 310 315 320
 Lys His Glu Pro Val Phe Tyr Val Arg Thr Glu Arg Ser Ser Ser Gly
 325 330 335
 Arg Ser Ile Thr Thr Val Val Asn Gly Val Thr Tyr Lys Asn Leu Arg
 340 345 350
 Phe Phe Ile His Pro Tyr Asn Phe Val Ser Ser Lys Thr Gln Arg Ile
 355 360 365
 Met

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTAAAACGAC GGCCAGT

17

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCCAAGCTTG GATGAT

16

- 89 -

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATCTTCGCGA ATTCACTGGC CGTCGTTTTC C

31

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAATTCGCGA AGAT

14

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCATCCAAG CTGGCACTG GCCGTCGTTT TAC

33

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 81 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTAAACGAC GGCCAGTGAA TTCGCGAAGA TNNNNNNNNN NNNNNNNNAT CATCCAAGCT

60

TGGCACTGGC CGTCGTTTTC C

81

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTAAACGAC GGCCAGTGCC AAGCTTGGAT GATNNNNNNN NNNNNNNNNN ATCTTCGCGA 60
ATTCAGTGGC CGTCGTTTTA C 81

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

```
(ix) FEATURE:
      (A) NAME/KEY: CDS
      (B) LOCATION: join(26..148, 190..207, 244..270)
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

[illegible]

- 91 -

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

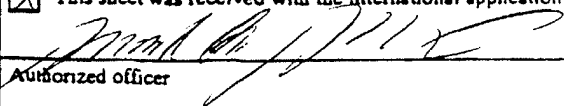
(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met	Thr	Met	Ile	Thr	Pro	Ser	Ser	Lys	Leu	Thr	Leu	Thr	Lys	Gly	Asn
1				5				10						15	
Lys	Ser	Trp	Tyr	Arg	Gly	Pro	Pro	Ser	Arg	Ser	Thr	Val	Ser	Ile	Ser
		20						25					30		
Leu	Ile	Asn	His	Leu	Tyr	Asn	Lys	Arg	Met	Leu	Ser	Tyr	Tyr	Thr	Met
		35					40					45			
Ser	Phe	Arg	Thr	Leu	Glu	Leu	Phe								
	50						55								

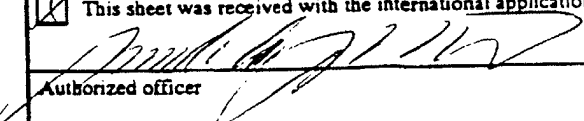
INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>79</u> , line <u>13</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 UNITED STATES OF AMERICA	
Date of deposit November 6, 1992	Accession Number A.T.C.C. 75354
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

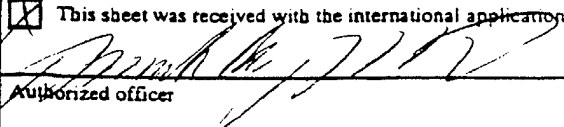
INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>79</u> , line <u>10</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>12301 Parklawn Drive</u> <u>Rockville, Maryland 20852</u> <u>UNITED STATES OF AMERICA</u>	
Date of deposit <u>January 21, 1993</u>	Accession Number <u>A.T.C.C. 75399</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<p>"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."</p>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
<div>For receiving Office use only</div> <div><input checked="" type="checkbox"/> This sheet was received with the international application</div> <div> Authorized officer</div>	<div>For International Bureau use only</div> <div><input type="checkbox"/> This sheet was received by the International Bureau on:</div> <div>Authorized officer</div>

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>31</u> . line <u>25</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 UNITED STATES OF AMERICA	
Date of deposit June 30, 1994	Accession Number A.T.C.C. 69341
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<p>"In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 23(4) EPC)."</p>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

WE CLAIM:

1. A purified and isolated polynucleotide encoding a CviJI polypeptide or a variant thereof possessing activity characteristic of CviJI, said polynucleotide comprising a polynucleotide as set out in SEQ ID NO: 2.
2. The polynucleotide of claim 1 which is a DNA.
3. The DNA of claim 2 which is a viral genomic DNA sequence or a biological replica thereof.
4. The DNA of claim 2 which is a wholly or partially chemically synthesized DNA or biological replica thereof.
5. A purified isolated DNA encoding a polypeptide according to claim 1 by means of degenerate codons.
6. A vector comprising a DNA according to claim 2.
7. The vector of claim 6 which is the plasmid pCJH1.4 (ATCC Accession No. 69341).
8. A host cell stably transformed or transfected with a DNA according to claim 2 in a manner allowing the expression in said host cell of a CviJI polypeptide or a variant thereof possessing a sequence specificity characteristic of CviJI.
9. The host cell according to claim 8, wherein said host cell is *E. coli*.

10. A method for producing a CviJI polypeptide or a variant thereof possessing biological activity specific to CviJI, said method comprising the steps of:

- a) growing a transformed host cell containing a vector according to claim 6 in a suitable nutrient medium; and
- b) isolating the CviJI polypeptide or variant thereof from said host cell.

11. The method of claim 10 wherein said host cell is *E. coli*.

12. A recombinant CviJI polypeptide.

13. A polypeptide produced by the method of claim 10.

14. A method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is cleaved at a dinucleotide sequence selected from the group consisting of PyGCPy, PuGCPy, PuGCPu, and wherein Pu = purine and Py = pyrimidine.

15. A method for restriction endonuclease digestion of DNA comprising the step of digesting DNA with a restriction endonuclease reagent under conditions wherein said DNA is digested at 11 of 16 possible dinucleotide sequences and wherein said dinucleotide sequences are selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

16. The method according to claim 14 wherein said restriction endonuclease reagent comprises CviI I.

17. A restriction endonuclease reagent, said restriction endonuclease reagent comprising in combination, Taq I and Hpa II (CGase I), said reagent capable of digesting DNA at 11 of 16 possible dinucleotide sequences, said sequences selected from the group consisting of PuCGPu, PuCGPy, PyCGPy and PyCGPu, and wherein Pu = purine and Py = pyrimidine.

18. The method according to claim 15 wherein said restriction endonuclease reagent is selected from the group consisting of Aci I and CGase I.

19. The method according to claim 16 wherein said digestion of DNA is a partial digestion and wherein said digestion generates quasi-random fragments of DNA without apparent site preference as seen on a 1-2 wt. % agarose gel.

20. The method according to claim 18 wherein said digestion of DNA is a partial digestion and wherein said digestion generates quasi-random fragments of DNA without apparent site preference as seen on a 1-2 wt. % agarose gel.

21. The method according to claims 16 or 18 wherein said digestion is complete, and wherein said digestion generates DNA fragments from about 20 base pairs in length to about 200 base pairs in length and wherein said fragments have an average length of about 20 to about 60 nucleotides.

22. The method according to claims 19 or 20 wherein said quasi-random fragments are from about 100 basepairs to about 10,000 base pairs in length.

23. A method for shotgun cloning and sequencing DNA, comprising the steps of:

- a) partially digesting DNA according to claims 19 or 20;
- b) ligating said partially digested DNA into a linearized cloning vector thereby creating a recombinant vector;
- c) introducing said recombinant vector into a host cell;
- d) selecting said host cell for the presence of said recombinant vector;
- e) growing and amplifying said host cell containing said recombinant vector;
- f) isolating and purifying said recombinant vector from said grown and amplified host cells; and
- g) sequencing said DNA contained in said recombinant vector.

24. The method according to claim 23 wherein said restriction endonuclease reagent comprises CviI I.

25. The method according to claim 23 wherein said restriction endonuclease reagent comprises CGase I.

26. The method according to claim 23 wherein said quasi-random fragments are from about 100 base pairs to about 10,000 base pairs in length.

27. The method according to claim 23 wherein said quasi-random fragments are from about 500 bp to about 2,000 bp in length.

28. The method according to claim 23 wherein said cloning vector is selected from the group consisting of plasmids, phage, and cosmids.

29. The method according to claim 28 wherein said plasmid is pUC19.

30. The method according to claim 28 wherein said bacteriophage is λ .

31. The method according to claim 28 wherein said bacteriophage is M13.

32. The method according to claim 23 wherein said host cell is a bacteria.

33. The method according to claim 32 wherein said host cell is *E. coli*.

34. The method according to claim 23 wherein said sequencing is dideoxy sequencing.

35. A kit for the shotgun cloning of DNA, said kit comprising in association:

- a) a restriction endonuclease reagent, according to claims 16 or 18;
- b) a restriction endonuclease buffer;
- c) ligation buffer; and
- d) T4 DNA ligase.

36. The kit of claim 35 further comprising in association:

- e) competent host bacteria;
- f) chromatography matrix said matrix useful for the size selection of restriction endonuclease digested DNA;
- g) spin filters, said spin filters useful for the size selection of restriction endonuclease digested DNA;
- h) a cloning vector;
- i) positive control DNA useful in the monitoring of the efficiency of the said shotgun cloning; and
- j) molecular size marker DNA.

37. The kit according to claim 35 wherein said restriction endonuclease reagent comprises CviI I.

38. The kit according to claim 37 wherein said restriction endonuclease buffer endonuclease buffer is CviI I^{**} buffer.

39. The kit according to claim 35 wherein said restriction endonuclease reagent comprises CGase I.

40. The kit according to claim 39 wherein said restriction endonuclease buffer is CGase I buffer.

41. The kit according to claim 36 wherein said competent host bacteria is competent *E. coli* DH5 α F'.

42. The kit according to claim 36 wherein said chromatography matrix is Sephacryl-S500.

43. The kit according to claim 36 wherein said cloning vector is M13 mp18.

44. A method for labeling DNA, the method comprising the steps of:

- a) digesting an aliquot of template DNA with a restriction endonuclease reagent according to claim 21 and wherein said digestion generates sequence-specific DNA fragments;
- b) mixing an aliquot of undigested template DNA with said sequence-specific DNA fragments, denaturing said mixture of template DNA and sequence-specific DNA fragments thereby generating denatured template DNA and oligonucleotide primers.
- c) annealing said primers to said denatured undigested template DNA to form a DNA-primer complex;
- d) performing an extension reaction from said primers in said DNA-primer complex using a DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label.

45. The method according to claim 44 wherein said restriction endonuclease reagent comprises CviI I.

46. The method according to claim 44 wherein said restriction endonuclease reagent comprises CGase I.

47. The method according to claim 44 wherein said extension reaction is performed by a DNA polymerase.

48. The method according to claim 47 wherein said DNA polymerase is *Thermus flavus* DNA polymerase.

49. The method according to claim 44 wherein the one or more nucleotide triphosphates are selected from the group consisting of dATP, dCTP, dGTP, dUTP and dTTP.

50. The method according to claim 44 wherein said labeled nucleotide triphosphate is selected from the group consisting of ^{32}P -labeled nucleotide triphosphates and ^{33}P -labeled nucleotide triphosphates.

51. The method according to claim 44 wherein said labeled nucleotide triphosphate is selected from the group consisting of biotin-labeled nucleotide triphosphates, fluorescein-labeled nucleotide triphosphates, dinitrophenol-labeled nucleotide triphosphates, and digoxigenin-labeled nucleotide triphosphates.

52. A method for thermal cycle labeling DNA comprising the steps of:

- a) digesting an aliquot of template DNA with a restriction endonuclease reagent according to claim 21 and wherein said digestion generates sequence-specific DNA fragments;
- b) mixing an aliquot of undigested template DNA with said sequence-specific DNA fragments, denaturing said mixture of template DNA and said DNA fragments thereby generating denatured template DNA and oligonucleotide primers;
- c) annealing said primers to said denatured undigested template DNA to form a DNA-primer complex;
- d) performing an extension reaction from said primers in said DNA-primer complex using a DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label.
- e) heat-denaturing said labeled extension products;
- f) reannealing said excess primers with said template DNA and with said extension products;
- g) performing at least one additional extension reaction from said DNA-primer complex using a DNA polymerase.

53. The method according to claim 52 wherein said restriction endonuclease reagent comprises CviI I.

54. The method according to claim 52 wherein said restriction endonuclease comprises CGase I.

55. The method according to claim 52 wherein said DNA polymerase is a heat stable DNA polymerase.

56. The method according to claim 55 wherein said heat-stable DNA polymerase is *Thermus flavus* DNA polymerase or a functional fragment thereof.

57. The method according to claim 52 wherein said extension products also serve as templates.

58. The method according to claim 52 wherein said label is selected from the group consisting of fluorescein, dinitrophenol, biotin, and digoxigenin.

59. The method according to claim 52 wherein said label is selected from the group consisting of ^{32}P , ^{33}P , ^3H , ^{14}C , and ^{35}S .

60. The method according to claim 52 wherein steps e)-g) are repeated up to 20 times.

61. A kit for labeling DNA, said kit comprising in association:
- a) a restriction endonuclease reagent, according to claims 16 or 18;
 - b) a restriction endonuclease buffer; and
 - c) a labeling buffer.

62. The kit according to claim 61 wherein said restriction endonuclease reagent comprises CviI I.

63. The kit according to claim 62 wherein said restriction endonuclease buffer is CviI I* restriction endonuclease buffer.

64. The kit according to claim 61 wherein said restriction endonuclease reagent is selected from the group consisting of CGase I and Aci I.

65. The kit according to claim 64 wherein said restriction endonuclease buffer is CGase I buffer.

66. The kit of claim 64 further comprising:

- d) a concentrated mixture of 1 or more nucleotide triphosphates;
- e) a DNA polymerase;
- f) control DNA, said control DNA being useful for monitoring the efficiency of labeling.

67. The kit according to claim 66 wherein said nucleotide mixture is an equimolar mixture of one or more nucleotides selected from the group consisting of dCTP, dTTP, dATP, and dGTP.

68. The kit according to claim 66 additionally comprising a labeled nucleotide selected from the group consisting of biotin-11-dUTP, digoxigenin-11-dUTP and fluorescein-11-dUTP.

69. The kit according to claim 66 additionally comprising a labeled nucleotide selected from the group consisting of ^{32}P -labeled nucleotides, ^{33}P -labeled nucleotides, ^{14}C -labeled nucleotides, ^{35}S -labeled nucleotides, and ^3H -labeled nucleotides.

70. The kit according to claim 66 wherein said DNA polymerase is the Klenow fragment of DNA polymerase 1.

71. The kit according to claim 66 wherein said DNA polymerase is a thermostable DNA polymerase.

72. The kit according to claim 66 wherein said thermostable DNA polymerase is *Thermus flavus* DNA polymerase.

73. A method for universal thermal cycle labelling DNA comprising the steps of:

- a) mixing an aliquot of template DNA with a holo-enzyme of a thermostable DNA polymerase, whereby the polymerase provides endogenously purified DNA primers;
- b) denaturing said mixture of template DNA and said endogenous DNA primers;
- c) annealing said mixture of denatured template DNA and said endogenous DNA primers to form a DNA-primer complex;
- d) performing an extension reaction from said endogenous DNA primers in said DNA-primer complex using said DNA polymerase in the presence of one or more nucleotide triphosphates and wherein at least one nucleotide triphosphate has a label;
- e) heat-denaturing said labeled extension products;
- f) reannealing said endogenous primers with said template DNA and with said extension products;
- g) performing at least one additional extension reaction from said DNA-primer complex using a DNA polymerase.

74. The method according to Claim 73 wherein said heat-stable DNA polymerase is *Thermus flavus* DNA polymerase or a functional fragment thereof.

75. The method according to claim 73 wherein said extension products also serve as templates.

76. The method according to claim 73 wherein said label is selected from the group consisting of fluorescein, dinitrophenol, biotin, and digoxigenin.

77. The method according to claim 73 wherein said label is selected from the group consisting of ^{32}P , ^{33}P , ^3H , ^{14}C , and ^{35}S .

78. The method according to claim 73 wherein steps e)-g) are repeated up to 20 times.

79. A kit for labeling DNA, said kit comprising in association:

- a) a holo-enzyme of a thermostable DNA polymerase;
- and
- b) a DNA polymerase buffer.

80. The kit of claim 79 further comprising:

- c) a concentrated mixture of 1 or more nucleotide triphosphates;
- d) control DNA, said control DNA being useful for monitoring the efficiency of labeling.

81. The kit according to claim 80 wherein said nucleotide mixture is an equimolar mixture of one or more nucleotides selected from the group consisting of dCTP, dTTP, dATP, and dGTP.

82. The kit according to claim 80 additionally comprising a labeled nucleotide selected from the group consisting of biotin-11-dUTP, digoxigenin-11-dUTP and fluorescein-11-dUTP.

83. The kit according to claim 80 additionally comprising a labeled nucleotide selected from the group consisting of ^{32}P -labeled nucleotides, ^{33}P -labeled nucleotides, ^{14}C -labeled nucleotides, ^{35}S -labeled nucleotides, and ^3H -labeled nucleotides.

84. The kit according to claim 80 wherein said thermostable DNA polymerase is *Thermus aquaticus* DNA polymerase.

85. The kit according to claim 80 wherein said thermostable DNA polymerase is *Thermus flavus* DNA polymerase.

86. A method for labeling of restriction-generated oligonucleotides, the method of comprising the steps of:

- a) digesting an aliquot of template DNA according to claim 21;
- b) heat denaturing said digested DNA thereby generating sequence-specific oligonucleotides; and
- c) labeling said sequence-specific oligonucleotides with a label capable of detection.

87. The method according to claim 86 wherein said restriction-generated oligonucleotides are labeled on the 5' end.

88. The method according to claim 86 wherein said restriction-generated oligonucleotides are labeled on the 3' end.

89. The method according to claim 86 wherein the label is radioactive.

90. The method according to claim 86 wherein the label is non-radioactive.

91. A method for anonymous primer cloning, the method comprising the steps of:

- a) digesting an aliquot of template DNA according to claim 21 thereby generating anonymous DNA fragments;
- b) digesting a plasmid cloning vector with a restriction endonuclease thereby creating a cloning site for insertion of said anonymous DNA fragments;
- c) ligating the anonymous DNA fragments of step a) into the cloning site of step b) thereby creating recombinant plasmids;
- d) transforming competent bacteria with the recombinant plasmids;
- e) selecting transformed colonies;
- f) purifying the recombinant plasmids from said transformed bacteria;
- g) digesting the recombinant plasmid with a restriction endonuclease said restriction endonuclease being capable of cutting said recombinant plasmid at a site, said site lying within the cloned anonymous DNA fragment;
- h) annealing one or more extension primers to the digested recombinant plasmid, said extension primers being complementary to plasmid sequences flanking the anonymous primer;
- i) extending the extension primer in a template-dependent fashion in the presence of one or more nucleotide triphosphates and a DNA polymerase; and
- j) denaturing the said hybridized extended primer.

92. The method according to claim 91 wherein said restriction endonuclease reagent comprises CviI I.

93. The method according to claim 91 wherein said restriction endonuclease reagent comprises CGase I.

94. The method according to claim 91 wherein said plasmid cloning vector is pFEM.

95. The method according to claim 94 wherein the restriction endonuclease of step b) is Eco RV.

96. The method according to claim 91 wherein said extension primer has a label capable of detection.

97. A kit for anonymous primer cloning comprising in association:

- a) a restriction endonuclease reagent, according to claims 16 or 18;
- b) a restriction endonuclease buffer;
- c) a cloning vector;
- d) competent bacteria;
- e) one or more extension primers said extension primers being complementary to plasmid sequences flanking said anonymous primers; and
- f) a DNA polymerase reagent.

98. The kit according to claim 97 wherein said restriction endonuclease reagent comprises CviI I.

99. The kit according to claim 98 wherein said restriction endonuclease buffer is CviI I* buffer.

100. The kit according to claim 97 wherein said restriction endonuclease reagent is selected from the group consisting of CGase I and Aci I.

101. The kit according to claim 100 wherein said restriction endonuclease buffer is CGase I buffer.

102. The kit according to claim 97 wherein said cloning vector is pFEM.

Figure 1

lacZ'
TAACAATTTCACACAGGAAACAGCT ATG ACC ATG ATT ACG CCA AGC TCG AAA TTA
M T M I T P S S K L
XhoI
ACC CTC ACT AAA GGG AAC AAA AGC TGG TAC CGG GGC CCC CCC TCG AGG TCG
T L T K G N K S W Y R G P P S R S
ACG GTA TCG ATA AGC TTG ATA AAC CAT TTA TAC AAT AAG CGT TGA TATAAGTTT
T V S I S L I N H L Y N K R
GTATATACGTCATTTTCGTTATATCAACAA ATG TTA TCA TAT TAT ACG TAA AACTGGCT
M L S Y Y T
M.CviJI
TAAAAAAAACGAGGTGTAACATA ATG TCT TTT CGC ACG TTA GAA CTA TTT ...
M S F R T L E L F

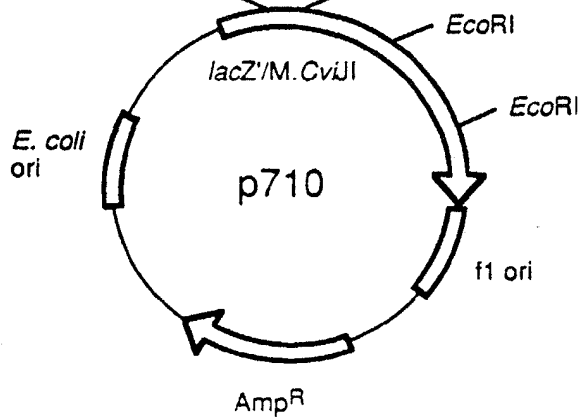


Figure 2

10 20 30 40 50 60 70 80

1 ATGCTTTTGC GCACGTTAGA ACTATTCCGC GGTATAGCTG GTATTTTACA TGGCTTCAGA GGTATATCTA CACGAGTTGC 80
8 ATTCGTAGAA ATTAATGAAG ACGCACAAAA ATCTTTGAAA ACAAAGTTT CAGATGCATC TGTATTTCAAT GACGTTACGA 160
16 AATTTACCAA ATCGGACTTC CCGAAGACAA TAGACATGAT TACGCGGGA TTCCCTTGCA CTGGGTTAG TATTCGAGGT 240
24 TTAGAACTG GATTGCAACA CAAGGAATTC GGTCTTTTC CCGATGTTG GCGAATCAGG GAAGAGTATA AACTTAAAA 320
32 AGTATTTTGC GAAACTCTCC ATATGTTGTC CCACACTTAC AATCTCGATG TCGTGTAAA AAAGATGAT GAAATTTGGT 400
40 ATTTCTGCAA GTGGGTAACT TGTGGGCAAT CAATTATAGG AGCCCATCAT CAACGCCACC GGTGGTTTTC TCTCGGATT 480
48 CGAAAGATT ATGAACCAGA AGAAATAATT GTATCTGTGA ATGTACAAA GTTCGACTGG GAAATAATG AACCCCGTG 560
56 TCAAGTAGAC AATAAGATT ACGAGAATTC AACTTTTGT CCGTGGCAG CATATTTCCG GGTCCCGGAC CAGATCAGAT 640
64 ATGCTTTTCA CCGTCTATT ACAGGTGATT TTGAGTCATC GTGAAAACCT ACCTTGACAC CTGGGACAAAT AATTTGGCAC 720
72 GAACACAAAA AATGAAGAG AACTTAGCAT AAAGTCATAA ACGGTATTA TGAGAACGAT GTGTATTATT CTTTTCAAG 800
80 GAAAGAAGTT CATCGCGTTC CTCTAAATAT ATCCGTGAAA CCAGCTGATA TTCCGGAGAA ACATAACGGA AAAACACTCG 880
88 TAGATCGGGA AATGATCAAG AAATATTGGT GCACACCATG TGTAGTTAT GGCACGTGCTA CTGCTGGATG CAATGTTCTG 960
96 ACAGACCGTC AGTCACATGC ACTTCTTACA CAAGTCAGGT TTTCATATAG GGTGTATGT GGACGACATT TGTCTGTAT 1040
104 ATGGTGTGCA TGGTTGATGG GGTATGACCA AGAATATCTT GGTATTTTGG TTCAATATGA TTAATAATTT TTGATACACT 1120
112 AAATGGATAT AAGAAGAAAA CGTTTTACAA TAGAAGGGGG TAAACGTATA ATACTCGAAA AAAAGAGACT TGAAGAGAAA 1200
120 AAAAGAATTG CGGAAGAGAA AAAAAGAAAT GCACCTTATG AAAACAACAG AATTGCGGAA GAGAAAAAAA GAATTTGCGGA 1280
128 AAGAAGAAAA CGATTGCGAC TTGAAGAGAA AAAACGAATT CGCGAAGAAA AAAAACGAAT CGCGGAAGAG AAAAACCGAA 1360
136 TCGTGGGAAGA GAAAAAAGA CTTGCACCTA TAGAAAAACA ACGAATTCGG GAAGAGAAAA TTGGCTGGGG GAGAAAAAAT 1440
144 AGAAAGAGGA TCTCTACAAA TGCAACAAAA CATGAAGAG AAATTTGTCAA AGTTATAAT TCAATTGCTC TCGGACCCCG 1520
152 TACTTTTGTG TCGTAGATA TAAAGGTAA TAAATCCAGA GAAATCCACA ACGTTGTAA AGTTAGACAA TTACAAGGCA 1600
160 GTAAAGCGAA ATCCCGGACC GCGTATGTTG ATAGAGAAAT TAACAAACCT AAAGCGGATA TAGGACCGGT AGACATAAC 1680
168 GTTACAGGA AAAGAATTAG AAGAAGTTCT ATCGTTCAAG AGATAACA ATATCTAAAA ATTTCTGGAA AGAACCTCAA 1760
176 CTGCTAATAA GACCGTATGG TCTGCTATCA AGTCAAAATTT GTTCAAAAA CAAGCAATAT TCGGATTTGA TTACGGTAAG 1840
184 AAACAGGAA GGGACAATGT AGACATCATA GGTCAAGGAC GACCAATTAT AACAAAAAGA GGTTCCTAT TATATCTTAC 1920
192 ATTCACGTGT TTTAGCGCAT TAAATGGGCA CTTGGAGAAT TTACTGGGA AACATGAACC CGTTTCTAT GTAAGAACAG 2000
200 AACCGAGTAG TAGCGGAGAA AGTATAACAA CTGCTGTCAA TGGTGTCACT TATAAAAAAT TAAGATTCTT TATACATCCA 2080
208 TACAACCTTT TTTCTTCAA AACACAACGT ATTATGTAGG ACCATTTTCC CGAGAGACTT TGTTGACCGG GTACTAAAAA 2160
216 ATGGTCACCA TATTTGTCTA AGATGCTCA TAGAAGCAGG TGCAAAACCT GACATCGTCA GTGTTGAGTA TACACATATA 2240
224 CATCTACATG TGGTGATATT TGTATAAACC GTAAATACCT ATATATACAA TACGTATCCC CCTAAAAAGC CTTAGATTTT 2320
232 TTAGTTGTAT ACTACTTTTG TATAAGACCT GTAAGTTACA AACTAAAAAT TTCAGCTTTG CTTTCGAAAC AAGCAATTAC 2400
240 CGCATGAGAA TAATATCCAT TATGGATGTT TTTCTGCTAAT AAAACGATAT TTCTACAGA AGTTTCTATG ATTAGTTCCG 2480
248 AAATATTGAG ATCATCGTCA CGTTTTTCTT TACCGTATTT TACTTTCTGT ATCGTCCGAC CAATAAAATC ATTCGTGTG 2560
256 AGTTTCTTCG GCAATTTGTC CGTGACACCA AATCTCTCAC AACACCTTTG ATGTCCATCC ATTTGCTAAC CTATCGGTAA 2640
264 TCCATGTTGT GTGTGTACGA CCACACCGTT ATAAGTATAA CACGTTAGT TGTGCTTAT ATCATAGCAT TCGAGAGCGG 2720
272 TGTGAACCTT TTCAGATCTA TTATTAATCG GATCTGATCC ATAAAGAGAA TCTTCATATT TACAATAAAA ATCATCCGAT 2800
280 ATGTTCTGCA CACGAACAAC ATTCGTCAAA TTTCTGTGAT AGCGAATCTC CATCTGTGAA TCATTAGAGA CTTGGCAGTA 2880
288 TATAACATTA TAATGTTTGA TATGATTATT ACGTTTCATA TCAACAAAA ACATATAAAC ACCATACAAA TATTAAAAA 2960
296 CGTTAGTATA TAATGGATAA CATTTGCAAT AGTATATTCA CTGAGTAAA AAATGGGCCA GAAGCTTGTG TGAAGATGAT 3040
304 GCTCATTTGA AGAGGTAGCA ATATCAATGA TGTTTCCGAA CTTAAAAATG GAAATACACC ACTACATATT TCGAGCTCAT 3120
312 ATGGTAATGA TGTGTGTTTG AAGATGCTTA TTGACGCGAG TGCAAACTTT GATATCACAG ATATTTCTCG AGGAACACCA 3200
320 CTTCATCTCG CGGTTTTGAA TGGCCATGAC ATATGTGTAT AGCGTGTAGT TGTGCTTAT ATCATAGCAT TCGAGAGCGG 3280
328 TAATTTGGGA TGGATACCGT TACATTACGC GGGTTTTAAT GGTAATGATG CGATTTTGA GATGCTCATC GTTGTAAAGT 3360
336 AATATTGTTG CGTTATCAAT GATCGCGGTT GGACGGCGTT ACATTACCGG GCTTTTAAAT GTCATAGCAT GTGCGTCAAG 3440
344 ACGCTTATTG ATCGCGGTGC AAATCTTGAC ATCAGAGATA TTTCCGGATG TACACCATT CATCGTGGG TTTATAATGA 3520
352 CCACGATGCA TGTGTGAAGA TACTCGTAGA AGCAGGTGCA ACTCTTACG TCATTGATGA TACTGAGTGG GTGCCGTTAC 3600
360 ATTACCGCGG TTTAATGGT AATGATGCGA TTTTGAGGAT GCTCATTTGA CGAGGTGCAG ATATTGATAT ATCTAATAT 3680
368 TGTGATTGGA CGGCGTTACA TTACCGCGCT CGAAATGGAC ACGATGTGTG TATAAAAAA CTCATCGAAG CAGGTGGTAA 3760
376 CATCAACGCC GTCAACAAAT CGGGGGATAC ACCACTAGAT ATTGCAGCAT GTCATGACAT TGCAGTATGT GCGAGCTGA 3840
384 TAGTCAATTA GATCGTTTCG GAGCGGCGGT TGGCTCCGAG TGAGTTGTGT GTCATACCAC CAACGCTCGC TGCATTAGGT 4000
400 GATGTGTTGC GAACGACGAT GCGGCTTCAT GGGCGATCGG AAGCTGCAAA GATCACAGCG CATCTTCCTG TGGGTGCAAG 4080
408 GGATACCTTA CGAACTACTG CGTTGTGTTT GAACCGAACA ATTTCCGAGA GATCTCGTTG ATAGTGTATT AATTGAATGC 4160
416 GTGTAAAGTT ACGCTATTTT TTTCCAAAAA GGGTTTGAT GAAATACAA ACATCTTTT GTAGATCTTT TACCATTAGT 4240
424 TAGTAGTATC TAAATTCAAC CCTTTGAAC CATCGCCATT AACAGACAGA GCGTATGAAC CGTTTTGTGC CAATTTCAAC 4320
432 TGTATTCTGT CAATAGAGAC CATACGTACC TCCAAATTTA TTTACTTTAC CTACAGTATT ACCACTTCTT TTTTTCCTA 4400
440 TTTCAAAACGA TAGTAACCCA TTGACCTCTA GGAATTTTAA CCCAGAACCT GAAATTTGTA AGAATCTGTC TTAATTTCAA GTCCTTTTT 4480
448 AAAAGCATAC AACGATCTGT TAAGGTTATC CCCAGAACCT GAAATTTGTA AGAATCTGTC TTAATTTCAA GTCCTTTTT 4560
456 GATCTGTATA CATATCACTT GGTTCGAAAT GAAATCTGTA GTCCCAATTA GGTACGTTCC ACCAAGTTTA ATACGGGGTC 4640
464 TTTCCACCGA GACCGGACAT TTCAGCACGA GCTTTGTAAG AATGATATGA TGTGGTTTAA TCTCTATCAC CATCGTTCCA 4720
472 CTTTCTCTG AACCGAAGAC CATGCATCGT TATACCTGGT GCAACCTGTA CTAAATTTCT TATTTAGGT GCGGCTCCGG 4800
480 GTGGATTAA CCGAGATTTC TCAATCTTAA AATATGATAA CGATGTCCA ACAGTAGAAC CACTGGGTGG TATGGCAGTT 4880
488 GCTGGAAGGG AAGGTAAAA TTTAGGATAT TTCAATACAC CAACACCTTG AGGGTTTACT TGAATACTTC TGGGAGATTG 4960
496 TGGTGGTTTC GTGCAAGGTG GTTTCGTTGA AGGTGGTTTC GTGCAAGGTG GTTTCGTCGA AGGTGGTTTC GTGCAAGGTG 5040
504 GTTTCGTCGA AGGTGGTTTC GTGCAAGGTG GTTTCGTCGA AGGTGGTTTC GTGCAAGGTG GTTTCGTCGA AGGTGGTTTC 5120
512 GTGCAAGGTG GTTTCGTCGA AGGTGGTTTC GTGCAAGGTG GTTTCGTCGA AGGTGGTTTC GTGCAAGGTG GTTTCGTCGA 5200
520 CGGAAGTGG GCATGACCAT AATCCGTTAA ATTCGCCAT TCTCTTAATG ATGTACTCCA TAAAGAACC GGTCCGCAAT 5280
528 GCATCTTAT TGGTCTGTGA GTATCAGATA TACATACGAA AATAAGAGAA TCATTTTCCC TGCCAAATAA TTTACAGAT 5360
536 TTTCCCTTAC ATGACATTAT TTGTAATATA ATATTATAT AATTTTAAA AAACCTAACG CTATTTTAAA TTATGTAATA 5440
544 CGTATTATAT CAATGCATCA TCTTAATCAT TTCTAACG TTAAGCGTAG CGAATTC

Figure 3

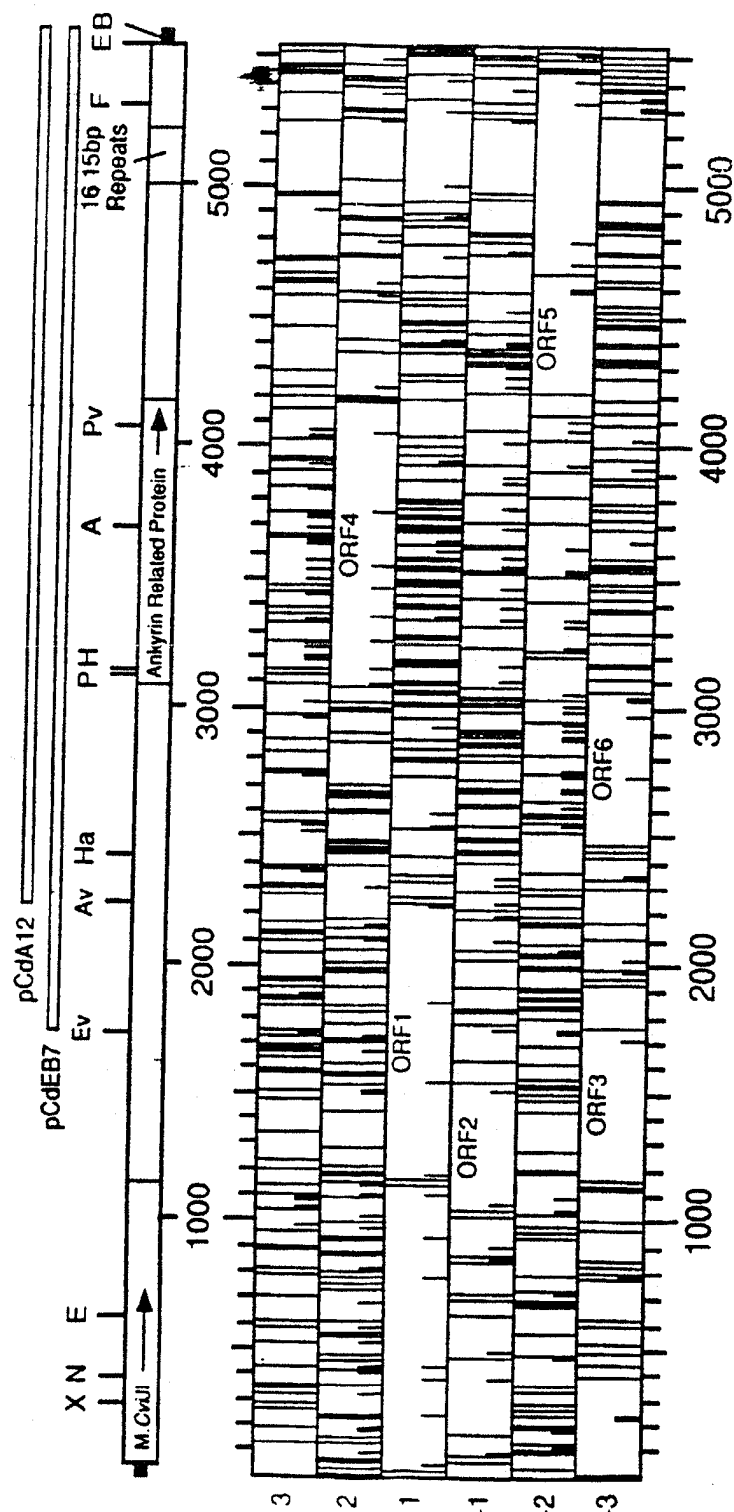


Figure 4

M.CvJl
1 CAA GAA TAT CTT GGT TAT TTG GTT CAA TAT GAT TAA AATATTTTGATACACTAA ATG GAT ATA
Q E Y L G Y L V Q Y D . m d i

68 AGA AGA AAA CGT TTT ACA ATA GAA GGG GCT AAA CGT ATA ATA CTC GAA AAA AAG AGA CTT
r r k r f t i . g a k r i i i . k k r i

128 GAA GAG AAA AAA AGA ATT GCG GAA GAG AAA AAA AGA ATT GCA CTT ATA GAA AAA CAA CGA
e . k k r i a . e . k k r i a i i . k q r

188 ATT GCG GAA GAG AAA AAA AGA ATT GCG GAA GAG AAA AAA CGA TTC GCA CTT GAA GAG AAA
i a e . k k r i a e . e k k r f a i . e e k

R.CvJl
248 AAA CGA ATT GCG GAA GAA AAA AAA CGA ATC GCG GAA GAG AAA AAA CGA ATC GTG GAA GAG
k r i a . e . k k r i a . e . k k r i v e f 3

308 AAA AAA AGA CTT GCA CTT ATA GAA AAA CAA CGA ATT GCG GAA GAG AAA ATT GCG TCG GGG
k k r l a l f k q r l a e e k i a s g 23

368 AGA AAA ATT AGA AAG AGG ATC TCT ACA AAT GCA ACA AAA CAT GAA AGA GAA TTT GTC AAA
R K I R K R I S T N A T K H E R E F V K 43

428 GTT ATA AAT TCA ATG TTC GTC GGA CCC GCT ACT TTT GTA TTC GTA GAT ATA AAA GGT AAT
V I N S M F V G P A T F V F V D I K G N 63

488 AAA TCC AGA GAA ATC CAC AAC GTT GTA AGA TTC AGA CAA TTA CAA GGC AGT AAA GCG AAA
K S R E I H N V V R F R Q L Q G S K A K 83

548 TCC CCG ACC GCG TAT GTT GAT AGA GAA TAT AAC AAA CCT AAA GCG GAT ATA GCA GCG GTA
S P T A Y V D R E Y N K P K A D I A A V 103

608 GAC ATA ACC GGT AAA GAT GTG GCA TGG ATA TCC CAT AAA GCA TCT GAA GGA TAT CAA CAA
D I T G K D V A W I S H K A S E G Y Q Q 123

668 TAT CTA AAA ATT TCT GGA AAG AAC CTC AAG TTC ACA GGA AAA GAA TTA GAA GAA GTT CTA
Y L K I S G K N L K F T G K E L E E V L 143

728 TCG TTC AAG AGA AAA GTA GTT AGT ATG GCA CCG GTA TCT AAA ATA TGG CCT GCT AAT AAG
S F K R K V V S M A P V S K I W P A N K 163

788 ACC GTA TGG TCT CCT ATC AAG TCA AAT TTG ATT AAA AAT CAA GCA ATA TTC GGA TTT GAT
T Y W S P I K S N L I K N Q A I F G F D 183

848 TAC GGT AAG AAA CCA GGA AGG GAC AAT GTA GAC ATC ATA GGT CAA GGA CGA CCA ATT ATA
Y G K K P G R D N V D I I G Q G R P I I 203

908 ACA AAA AGA GGT TCC ATA TTA TAT CTT ACA TTC ACT GGT TTT AGC GCA TTA AAT GGG CAC
T K R G S I L Y L T F T G F S A L N G H 223

968 TTG GAG AAT TTT ACT GGG AAA CAT GAA CCC GTT TTC TAT GTA AGA ACA GAA CGG AGT AGT
L E N F T G K H E P V F Y V R T E R S S 243

1028 AGC GGG AGA AGT ATA ACA ACT GTC GTC AAT GGT GTC ACT TAT AAA AAT TTA AGA TTC TTT
S G R S I T T V V N G V T Y K N L R F F 263

1088 ATA CAT CCA TAC AAC TTT GTT TCT TCA AAA ACA CAA CGT ATT ATG TAG GACCATTTTCCCGAG
I H P Y N F V S S K T Q R I M . 273

1152 AGACTTTGTTGACCGCGTACTAAAAAATGGTCACGATATTTGTCTAAAGATGCTCATAGAAGCAGGTGCAAACCTTGAC

Figure 5

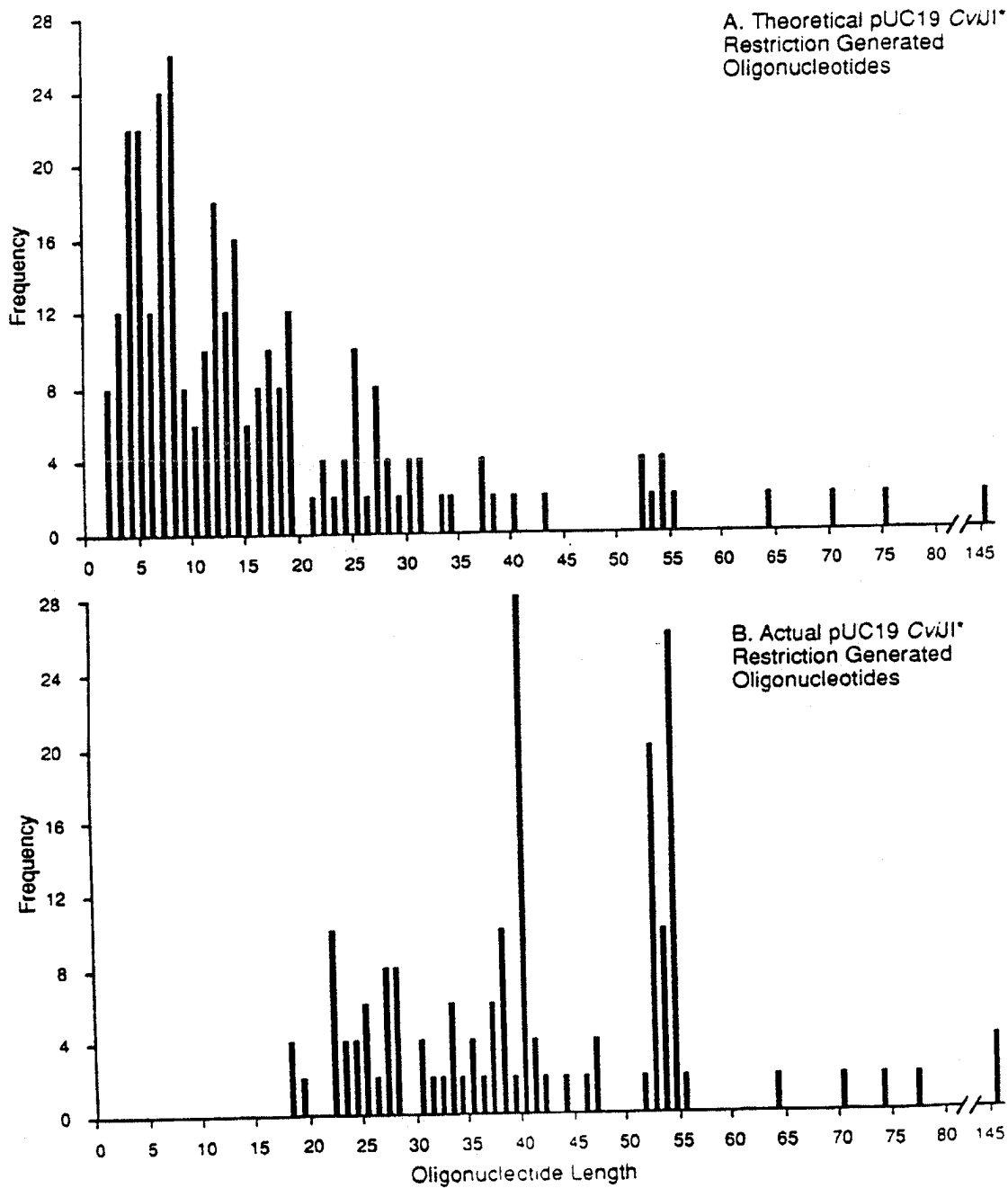


Figure 6

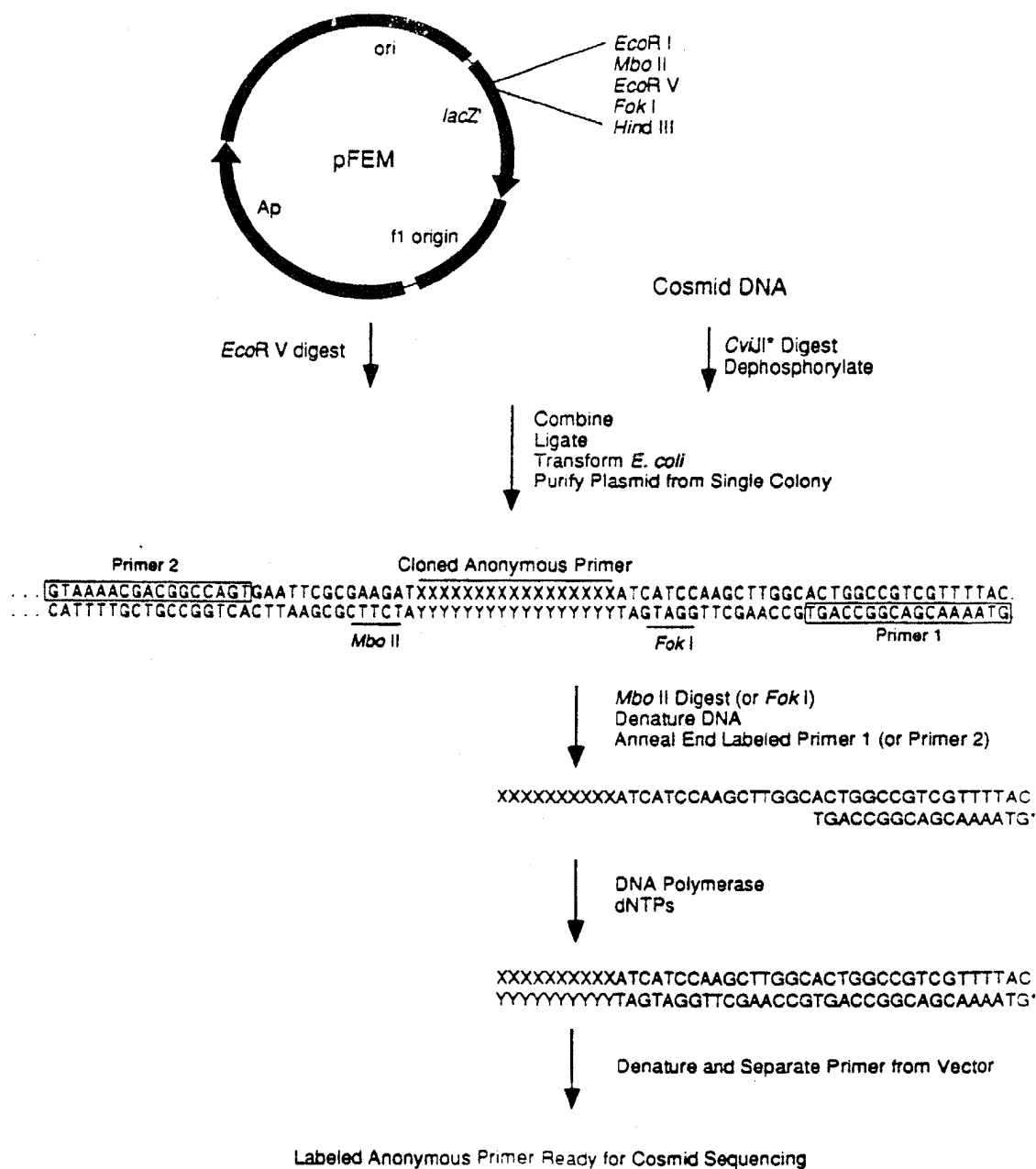


Figure 7

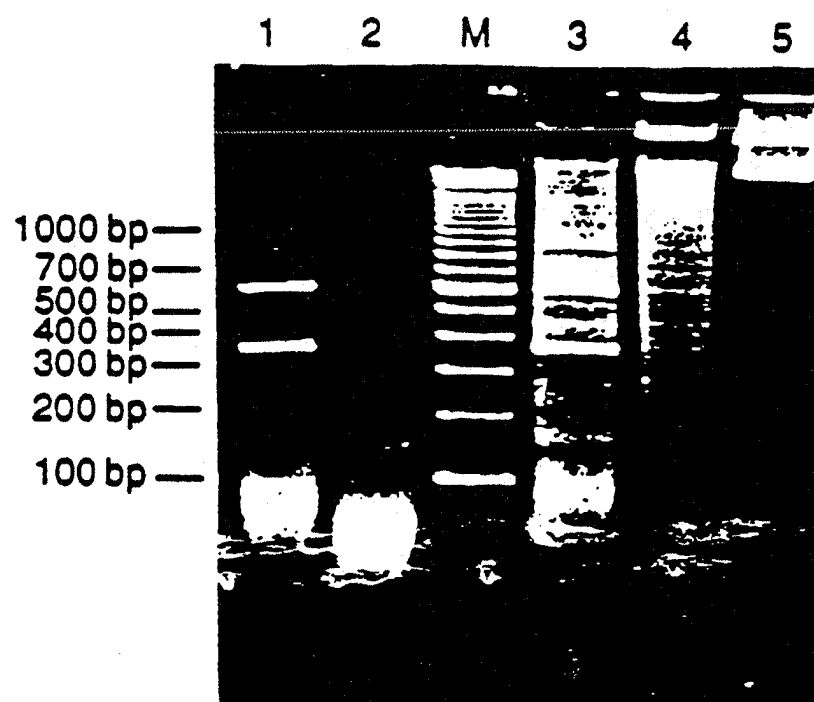


Figure 9

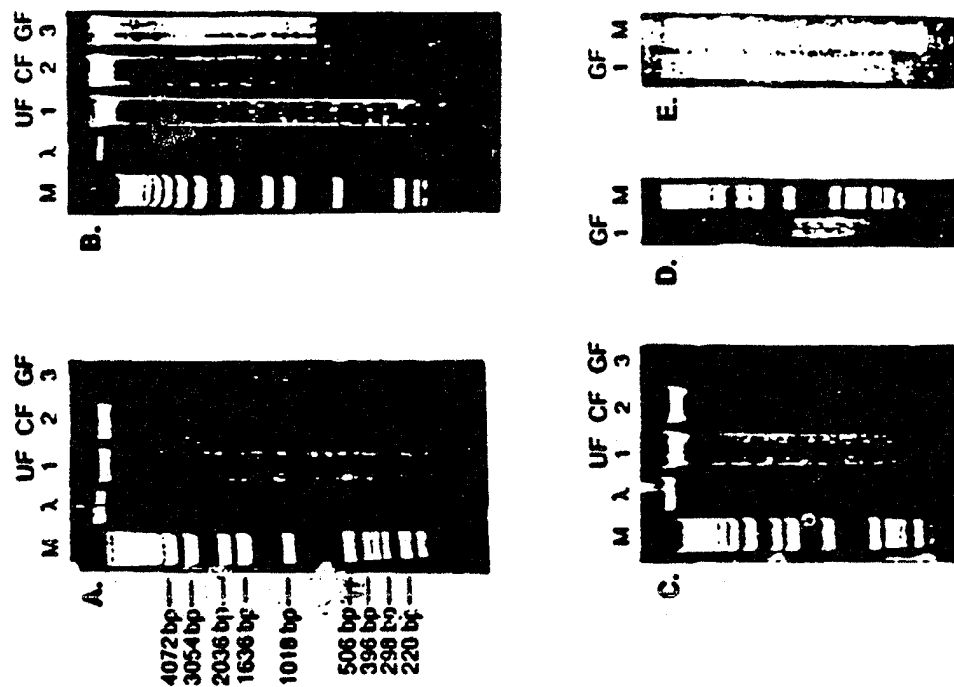


Figure 8

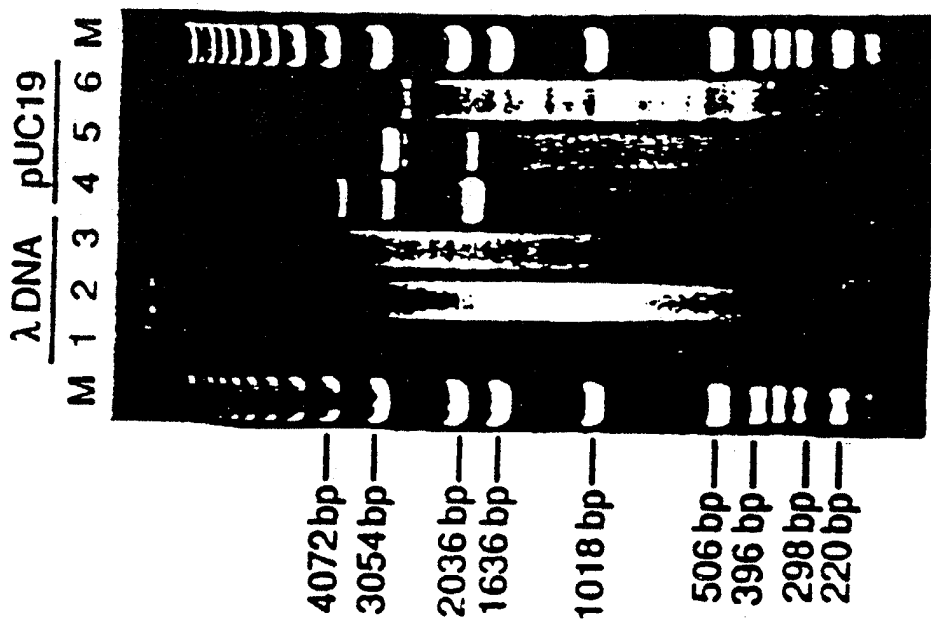


Figure 10

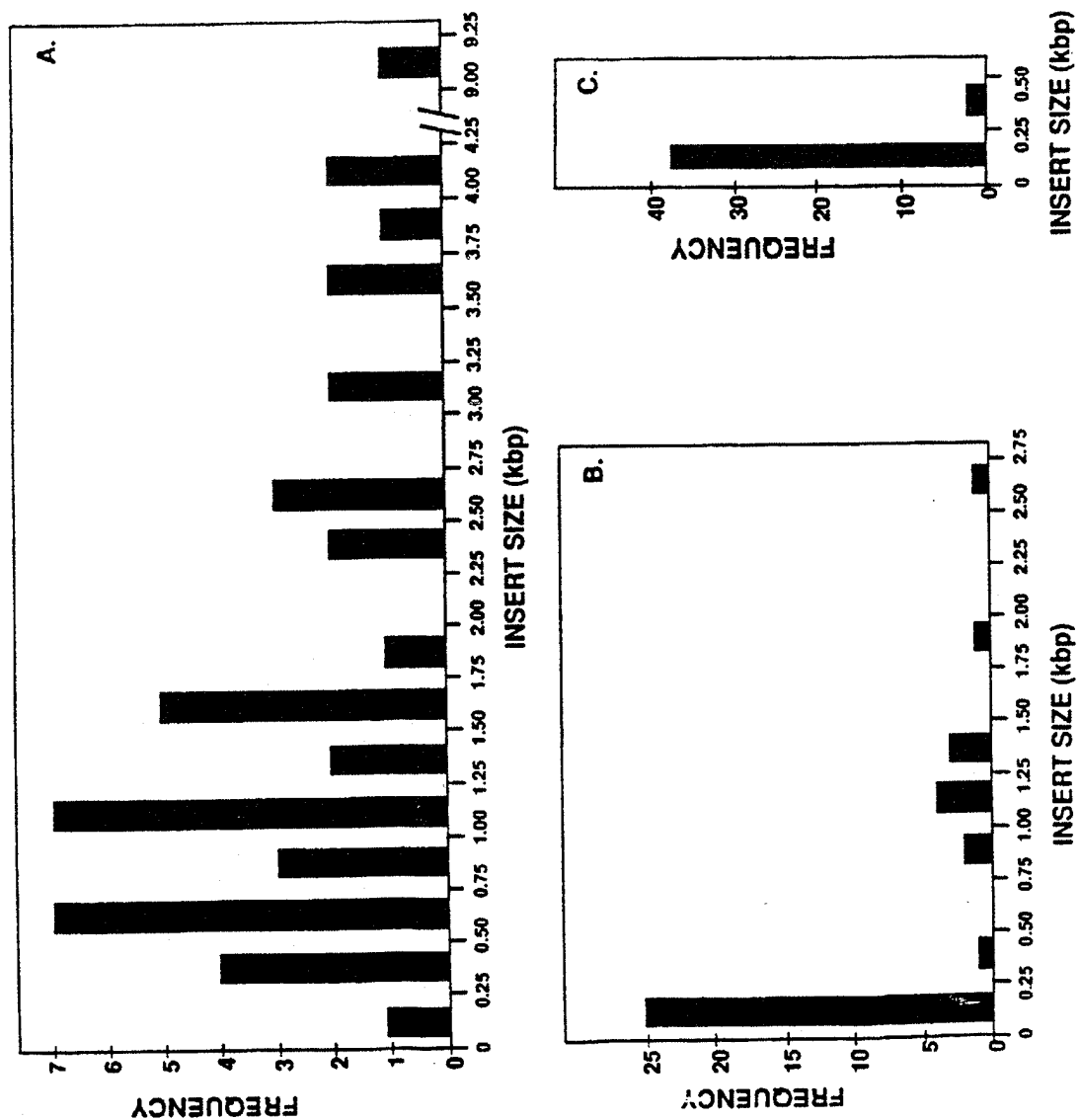


Figure 11

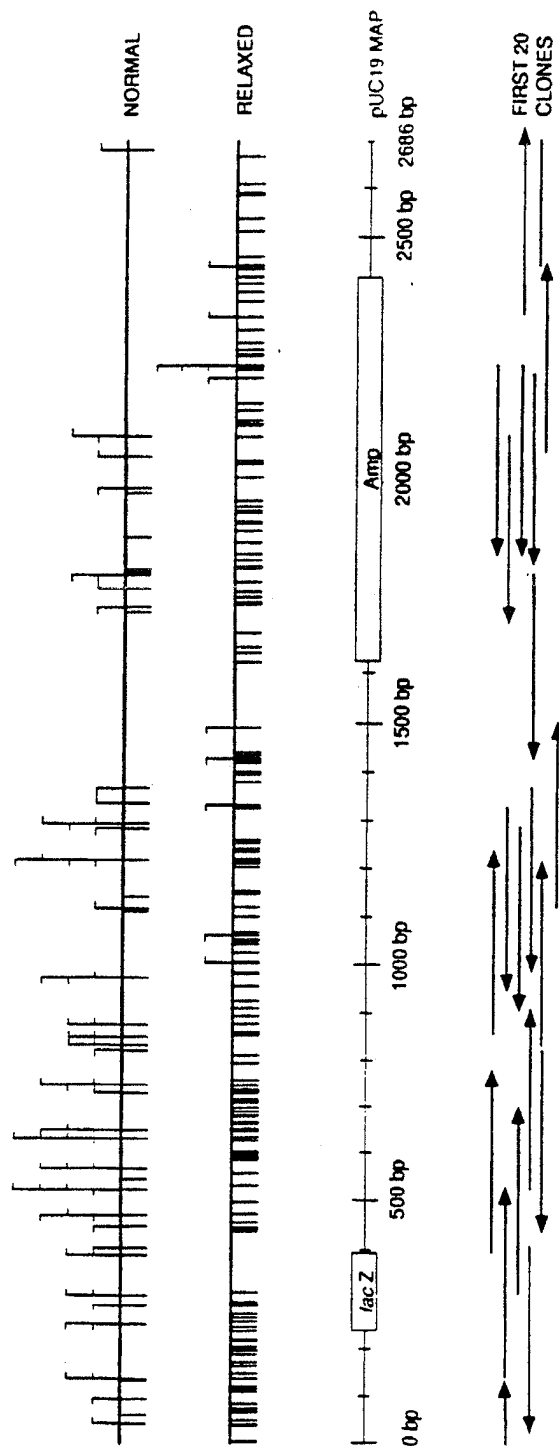
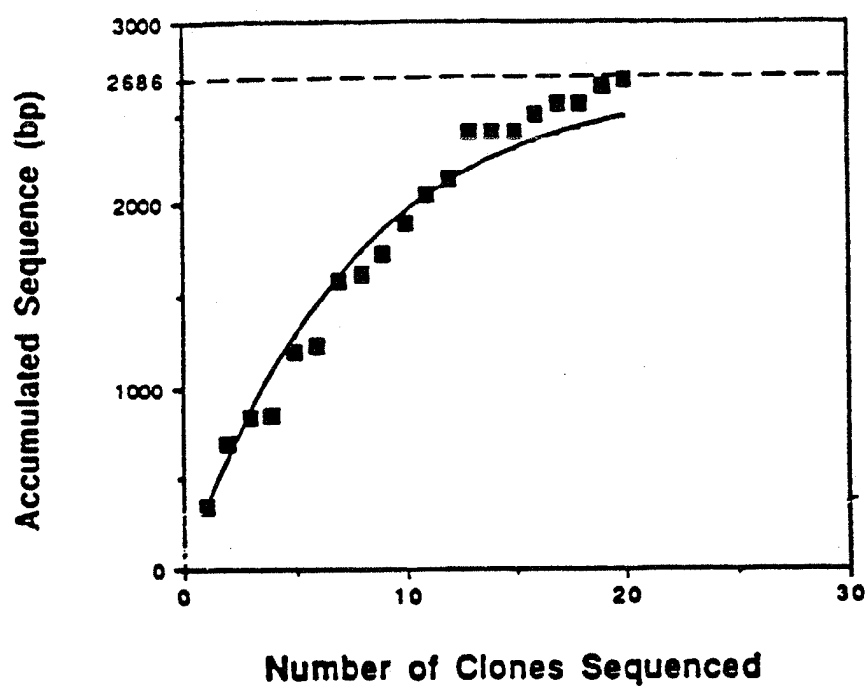


Figure 12



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/03246

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07H 21/02, 21/04; C07K 15/00; C12Q 1/68; C12P 19/34
US CL :435/6, 91.1, 91.53, 172.3; 530/350; 536/23.72

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.53, 172.3; 530/350; 536/23.72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CAS, WPI
search terms: CviJI, cloned restriction

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,053,330 (LUNNEN ET AL) 01 October 1991, column 2, lines 41-55.	1-11
Y	US, A, 5,075,232 (MORGAN) 24 December 1991, column 7, lines 32-43.	1-11
Y	Nucleic Acids Research, Volume 20, No. 7, issued 1992, Y. Zhang et al, "A single amino acid change restores DNA cytosine methyltransferase activity in a cloned chlorella virus pseudogene", pages 1637-1642, especially Figure 2.	1-11
Y	Virology, Volume 176, issued 1990, S. L. Shields et al, "Cloning and Sequencing the Cytosine Methyltransferase Gene M. CviJI from Chlorella Virus IL-3A", pages 16-24, especially Figure 3.	1-11



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

Z

document member of the same patent family

Date of the actual completion of the international search

29 JUNE 1994

Date of mailing of the international search report

JUL 13 1994

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DAVID SCHREIBER

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/03246

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-11, drawn to a polynucleotide encoding CviII, the vector carrying said polynucleotide, the transformed or transfected host carrying said vector, and method for producing a CviII polypeptide from said host, classified in Class 536, subclass 23.72, for example.
- II. Claims 12 and 13, drawn to the recombinant CviII polypeptide, classified in Class 530, subclass 350, for example.
- III. Claims 14, 16, 19, 21, and 22, drawn to a method for restriction endonuclease digestion using CviII, classified in Class 435, subclass 6, for example.
- IV. Claims 14, 15, 17, 18, and 20-22, drawn to CGase I restriction endonuclease and a method for using it in restriction endonuclease digestion, classified in Class 435, subclass 6, for example.
- V. Claims 23, 24, 26-38, and 41-43, drawn to a method for shotgun cloning after partial digestion using CviII, classified in Class 435, subclass 172.3.
- VI. Claims 23, 25-35, and 39-43, drawn to a method for shotgun cloning after partial digestion using CGase I, classified in Class 435, subclass 172.3.
- VII. Claims 44, 45, 47-53, and 55-63, drawn to a method of extension labeling of DNA and thermal cycle labeling using CviII, classified in Class 435, subclass 91.1, for example.
- VIII. Claims 44, 46-52, 54-61, and 64-72, drawn to a method of extension labeling of DNA and thermal cycle labeling using CGase I, classified in Class 435, subclass 91.1, for example.
- IX. Claims 73-85, drawn to a universal thermal cycle labeling of DNA, classified in Class 435, subclass 91.1, for example.
- X. Claims 86-90, drawn to a method of end labeling after CviII digestion, classified in Class 435, subclass 91.53.
- XI. Claims 86-90, drawn to a method of end labeling after CGase I digestion, classified in Class 435, subclass 91.53.
- XII. Claims 91, 92, and 94-99, drawn to a method for anonymous primer cloning after digestion with CviII, classified in Class 435, subclass 172.3, for example.
- XIII. Claims 91, 93, 94-97, and 100-102, drawn to a method for anonymous primer cloning after digestion with CGase I, classified in Class 435, subclass 172.3, for example.

Detailed Reasons for Lack of Unity

PCT Rule 13 recites the basic principle of unity of invention that an application should relate to only one invention or, if there is more than one invention, that applicant would have a right to include in a single application only those inventions which are so linked as to form a single general inventive concept. According to Rule 13, a group of inventions is linked to form a single inventive concept where there is a technical relationship among the inventions that involves at least one common or corresponding special technical feature that defines the contribution which each claimed invention, considered as a whole, makes over the prior art.

The thirteen inventions of this application consist of:

- 1) a polynucleotide encoding CviII, the vector comprising it, the transformed host carrying the vector, and a method of making the protein using the vector,
- 2) the recombinant peptide CviII,
- 3) a method for restriction endonuclease digestion using CviII,
- 4) CGase I restriction endonuclease and a method for using it in restriction endonuclease digestion,
- 5) a method for shotgun cloning after partial digestion using CviII,
- 6) a method for shotgun cloning after partial digestion using CGase I,
- 7) a method of extension labeling of DNA and thermal cycle labeling using CviII,
- 8) a method of extension labeling of DNA and thermal cycle labeling using CGase I,
- 9) a universal thermal cycle labeling of DNA,
- 10) a method of end labeling after CviII digestion,
- 11) a method of end labeling after CGase I digestion,
- 12) a method for anonymous primer cloning after digestion with CviII, and
- 13) a method for anonymous primer cloning after digestion with CGase I.

The thirteen inventions are not linked by a special technical feature within the meaning of PCT Rule 13 for the following reasons: Those claims drawn to CviII are not linked to those claims drawn to CGase I because there is no technical relationship among these inventions that involves at least one common or corresponding special technical feature.

The claims that involve the polynucleotide encoding CviII, the vector containing it, the host carrying the vector, and methods of making recombinant protein are not linked to the recombinant protein because the protein and polynucleotide share a technical relationship that involves a corresponding technical feature that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because cloning and expression of polynucleotides to make recombinant polypeptides are well known in the art. Accordingly, such does not

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/03246

constitute a special technical feature within the meaning of PCT Rule 13.2.

The methods for restriction endonuclease digestion, shotgun cloning and sequencing with CviJI, for extension and thermal cycle labeling with CviJI, for universal cycle labelling, for end labeling after CviJI digestion, and for anonymous primer cloning after CviJI digestion involve a corresponding technical feature, digestion with CviJI, that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because restriction endonuclease digestion, and shotgun cloning and sequencing, extension and thermal cycle labeling after rest, universal cycle labelling, end labeling, and anonymous primer cloning after restriction endonuclease digestion are well known in the art. In addition, CviJI is also known in the art. Accordingly, such does not constitute a special technical feature within the meaning of PCT Rule 13.2.

Similarly, the methods for restriction endonuclease digestion, shotgun cloning and sequencing with CGaseI, for extension and thermal cycle labeling with CGaseI, for universal cycle labelling, for end labeling after CGaseI digestion, and for anonymous primer cloning after CGaseI digestion involve a corresponding technical feature, digestion with CGaseI, that does not define the contribution which each claimed invention, considered as a whole, makes over the prior art because restriction endonuclease digestion, and shotgun cloning and sequencing, extension and thermal cycle labeling after rest, universal cycle labelling, end labeling, and anonymous primer cloning after restriction endonuclease digestion are well known in the art. Accordingly, such does not constitute a special technical feature within the meaning of PCT Rule 13.2.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



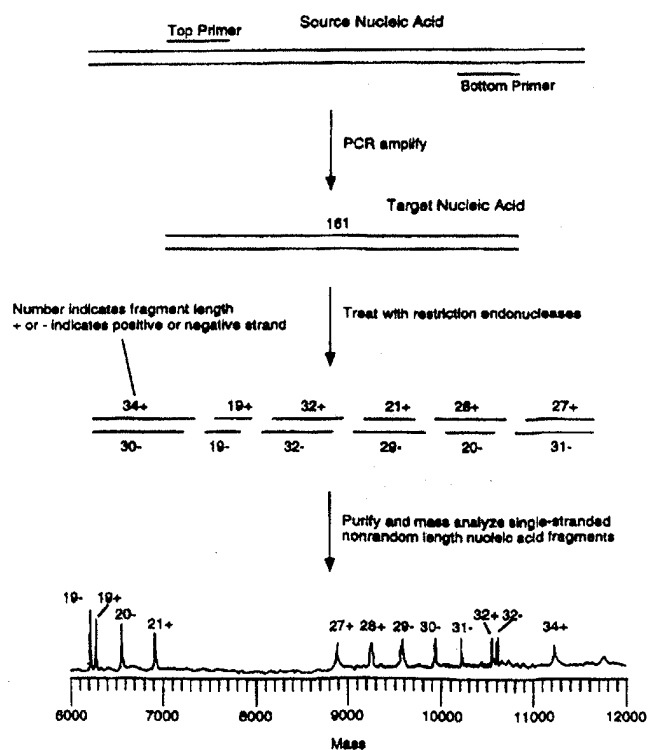
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68		A1	(11) International Publication Number: WO 97/33000
			(43) International Publication Date: 12 September 1997 (12.09.97)
(21) International Application Number: PCT/US97/03499			(81) Designated States: AU, CA, CN, IL, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
(22) International Filing Date: 4 March 1997 (04.03.97)			
(30) Priority Data: 60/012,752 4 March 1996 (04.03.96) US			
(71) Applicant: GENETRACE SYSTEMS, INC. [US/US]; 333 Ravenswood Avenue, Menlo Park, CA 94025 (US).			
(72) Inventors: MONFORTE, Joseph, A.; 50 Alamo Avenue, Berkeley, CA 94708 (US). SHALER, Thomas, A.; 1384 4th Avenue #3, San Francisco, CA 94122 (US). TAN, Yuping; 850 Coleman Avenue #16, Menlo Park, CA 94025 (US). BECKER, Christopher, H.; 106 Clover Lane, Menlo Park, CA 94025 (US).			
(74) Agents: NAKAMURA, Jackie, N. et al.; Cooley Godward L.L.P., Five Palo Alto Square, 3000 El Camino Real, Palo Alto, CA 94306-2155 (US).			Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: METHODS OF SCREENING NUCLEIC ACIDS USING MASS SPECTROMETRY

(57) Abstract

This invention relates to methods for screening nucleic acids for mutations by analyzing nonrandomly fragmented nucleic acids using mass spectrometric techniques and to procedures for improving mass resolution and mass accuracy of these methods of detecting mutations.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

METHODS OF SCREENING NUCLEIC ACIDS USING MASS SPECTROMETRY

5

ACKNOWLEDGEMENTS

This invention was supported in part by a Financial Assistance Award from the United States Department of Commerce, Advanced Technology Program, Cooperative Agreement #70NANB5H1029. The U.S. Government may have
10 rights in this invention.

TECHNICAL FIELD

This invention relates generally to methods for screening nucleic acids for mutations by analyzing fragmented nucleic acids using mass spectrometry.
15

INTRODUCTION

Approximately 4,000 human disorders are attributed to genetic causes. Hundreds of genes responsible for various disorders have been mapped, and
20 sequence information is being accumulated rapidly. A principal goal of the Human Genome Project is to find all genes associated with each disorder. The definitive diagnostic test for any specific genetic disease (or predisposition to disease) will be the identification of mutations in affected cells that result in alterations of gene function. Furthermore, response to specific medications may
25 depend on the presence of mutations. Developing DNA (or RNA) screening as a practical tool for medical diagnostics requires a method that is inexpensive, accurate, expeditious, and robust.

Genetic mutations can manifest themselves in several forms, such as point mutations where a single base is changed to one of the three other bases,
30 deletions where one or more bases are removed from a nucleic acid sequence and the bases flanking the deleted sequence are directly linked to each other, and

2.

insertions where new bases are inserted at a particular point in a nucleic acid sequence adding additional length to the overall sequence. Large insertions and deletions, often the result of chromosomal recombination and rearrangement events, can lead to partial or complete loss of a gene. Of these forms of mutation, in general the most difficult type of mutation to screen for and detect is the point mutation because it represents the smallest degree of molecular change. The term mutation encompasses all the above-listed types of differences from wild type nucleic acid sequence. Wild type is a standard or reference nucleotide sequence to which variations are compared. As defined, any variation from wild type is considered a mutation including naturally occurring sequence polymorphisms.

Although a number of genetic defects can be linked to a specific single point mutation within a gene, e.g. sickle cell anemia, many are caused by a wide spectrum of different mutations throughout the gene. A typical gene that might be screened using the methods described here could be anywhere from 1,000 to 100,000 bases in length, though smaller and larger genes do exist. Of that amount of DNA, only a fraction of the base pairs actually encode the protein. These discontinuous protein coding regions are called exons and the remainder of the gene is referred to as introns. Of these two types of regions, exons often contain the most important sequences to be screened. Several complex procedures have been developed for scanning genes in order to detect mutations, which are applicable to both exons and introns.

Gel Electrophoresis: Several of the procedures described below use some form of gel electrophoresis. Therefore it is worthwhile to briefly consider this separation technology before proceeding to the specific methods. In terms of current use, most of the methods to scan or screen genes employ slab or capillary gel electrophoresis for the separation and detection step in the assays. Gel electrophoresis of nucleic acids primarily provides relative size information based on mobility through the gel matrix. If calibration standards are employed, gel electrophoresis can be used to measure absolute and relative molecular weights of large biomolecules with some moderate degree of accuracy; even then typically the accuracy is only 5% to 10%. Also the molecular weight resolution is limited. In cases where two DNA fragments with identical number of base pairs can be

separated. using high concentration polyacrylamide gels, it is still not possible to identify which band on a gel corresponds to which DNA fragment without performing secondary labeling experiments. Gel electrophoresis techniques can only determine size and cannot provide any information about changes in base composition or sequence without performing more complex sequencing reactions. Gel-based techniques, for the most part, are dependent on labeling methods to visualize and discriminate between different nucleic acid fragments.

DNA Sequencing: The principal approach currently used to screen for genetic mutations is DNA sequencing. Sequencing reactions can be performed to screen the full genetic target base by base. This process, which can pinpoint the exact location and nature of mutation, requires labeling DNA, use of polyacrylamide gels, and a multiplicity of reactions to assess all bases over the length of a gene, all of which are slow and labor intensive procedures. [J. Bergh et al. "Complete Sequencing of the p53 Gene Provides Prognostic Information in Breast Cancer Patients, Particularly in Relation to Adjuvant Systemic Therapy and Radiotherapy," Nature Medicine 1, 1029 (1995)]

For DNA sequencing, nucleic acids comprising different exons or small clusters of exons are individually amplified, often using polymerase chain reaction (PCR). The amplifications are normally performed separately although some multiplexing of reactions is possible. The amplified nucleic acids typically range from one hundred to several thousand bases in length. Following amplification, the PCR products can serve as templates for standard dideoxy-based Sanger sequencing reactions. The four different sequencing reactions are run (or for fluorescence detection, one reaction with four different dye terminators) and then analyzed by polyacrylamide gel electrophoresis. Each sequencing run yields about 300 to 600 bases of sequence which typically must be read with at least a two to three-fold redundancy in order to assure accuracy. Using slab gel, the analysis process typically takes several hours.

SSCP: The single strand conformational polymorphism assay takes advantage of structural variation within DNA that results from mutation. The method involves folding the single-stranded form of a given nucleic acid sequence into a thermodynamically directed secondary and tertiary structure. In most cases, mutated sequences form different structures than the wild type sequence, thus permitting separation of mutated and wild type sequences by gel electrophoresis. Like sequencing, this assay is complicated by the need to label molecules and run polyacrylamide gels. In a typical case, mutations can be located within a general range of 50 to 200 base pairs, but the exact nature of the mutation cannot be identified. [M. Orita et al., "Detection of Polymorphisms of Human DNA by Gel Electrophoresis as Single-Stranded Conformation Polymorphisms," Proc. Natl. Acad. Sci. USA 86, 2766 (1989)]

DGGE: Like SSCP, denaturing gradient gel electrophoresis assays also differentiate based on structural variation, but require the use of gradient gels, which are difficult to prepare. The different thermodynamic stability of structures formed by the mutant sequence, as opposed to wild type, lead to differences in the temperature and/or pH at which the molecule will denature. DGGE mutation identification and localization properties are similar to those for SSCP though sensitivity is higher for DGGE because not all mutations cause the structural changes that the SSCP method depends upon for detection. [E.S. Abrams, S.E. Murdaugh & L.S. Lerman, "Comprehensive Detection of Single Base Changes in Human Genomic DNA Using Denaturing Gradient Gel Electrophoresis and a GC Clamp," Genomics 7, 463 (1990)]

EMC: Enzyme mismatch cleavage utilizes one or more enzymes that are capable of recognizing interruptions in base pairing within a double-stranded nucleic acid molecule, e.g. base-base mismatches, bulges, or internal loops. A given length of DNA or RNA is prepared in heterozygous form, with one strand composed of wild type nucleic acid and the other strand containing a potential mutation. At the specific site where the mutation forms a mismatch with the wild type sequence, a structural perturbation occurs. An enzyme such as T4 endonuclease VII, RuvC,

RNase A, or MutY, can recognize such a structural perturbation and can site-specifically cut the double-stranded nucleic acid, creating smaller molecules whose sizes indicate the presence and location of the mutation. As with the previously discussed methods, this approach as currently used, also requires labeling and gel electrophoresis. With this method, the site of mutation can be localized to within a few base pairs but the exact nature of the mutation cannot be determined. [R. Youil, B.W. Kemper & R.G.H. Cotton, "Screening for Mutations by Enzyme Mismatch Cleavage with T4 Endonuclease VII," Proc. Natl. Acad. Sci. USA 92, 87 (1995)]

CCM: A variation of EMC is to replace the enzymatic cleavage step with chemical cleavage. Chemical cleavage mismatch analysis involves the use of reagents such as osmium tetroxide to react with mismatched thymine residues or hydroxylamine to react with mismatched cytosine residues. Cleavage of the modified mismatched residues occurs when the modified bases are subsequently treated with piperidine or another oxidizing agent. The effectiveness of the method is similar to EMC. [J.A. Saleeba & R.G.H. Cotton, "Chemical Cleavage of Mismatch to Detect Mutations," Methods in Enzymology 217, 286 (1993)]

Hybridization Arrays: Several approaches to screening for mutations involve the probing of a target nucleic acid by an array of oligonucleotides that can differentiate between normal wild type nucleic acids and mutant nucleic acids. These arrays involve the performance of hundreds or thousands of hybridization reactions in parallel with different site-directed oligonucleotides and requires sophisticated and costly probe arrays. Hybridization arrays can identify the location and type of mutation in many, but not all cases. For example, semihomologous sequential insertions or targets with repeating sequences and/or repeating sequential motifs cannot be analyzed by hybridization. [A.C. Pease et al., "Light-Generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis," Proc. Natl. Acad. Sci. USA 91, 5022 (1994)]

Simple screens: For mutations localized within a given gene, such as the cystic fibrosis $\Delta F508$ deletion, it is also possible to perform a single PCR or ligase chain reaction (LCR) assay or simple hybridization assays tailored to these specific sites. PCR and LCR results are presently determined by the use of labeled molecules, where radioactive emissions, fluorescence, chemiluminescence or color changes are detected directly. These simple screens amount to a yes/no answer and do not directly identify the nature of the mutation, only whether or not a reaction took place. [P. Fang et al., "Simultaneous Analysis of Mutant and Normal Alleles for Multiple Cystic Fibrosis Mutations by the Ligase Chain Reaction," Human Mutation 6, 144 (1995)]

All of the methods in use today capable of screening broadly for genetic mutations suffer from technical complication and are labor and time intensive. There is a need for new methods that can provide cost effective and expeditious means for screening genetic material in an effort to reduce medical expenses. The inventions described here address these issues by developing novel, tailor-made processes that focus on the use of mass spectrometry as a genetic analysis tool. Mass spectrometry requires minute samples, provides extremely detailed information about the molecules being analyzed including high mass accuracy, and is easily automated.

The late 1980's saw the rise of two new mass spectrometric techniques for successfully measuring the masses of intact very large biomolecules, namely, matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (TOF MS) [K. Tanaka et al., "Protein and Polymer Analyses up to m/z 100,000 by Laser Ionization Time-of-flight Mass Spectrometry," Rapid Commun. Mass Spectrom. 2, 151-153 (1988); B. Spengler et al., "Laser Mass Analysis in Biology," Ber. Bunsenges. Phys. Chem. 93, 396-402 (1989)] and electrospray ionization (ESI) combined with a variety of mass analyzers [J. B. Fenn et al., Science 246, 64-71 (1989)]. Both of these two methods are suitable for genetic screening tests. The MALDI mass spectrometric technique can also be used with methods other than time-of-flight, for example, magnetic sector, Fourier-Transform, ion cyclotron resonance, quadrupole, and quadrupole trap. One of the advances in MALDI analysis of polynucleotides was the discovery of

3-hydroxypicolinic acid as an ideal matrix for mixed-base oligonucleotides. Wu, et al., *Rapid Comm'ns in Mass Spectrometry*, 7:142-146 (1993).

MALDI-TOF MS involves laser pulses focused on a small sample plate comprising analyte molecules (nucleic acids) embedded in either a solid or liquid matrix comprising a small, highly absorbing compound. The laser pulses transfer energy to the matrix causing a microscopic ablation and concomitant ionization of the analyte molecules, producing a gaseous plume of intact, charged nucleic acids in single-stranded form. If double-stranded nucleic acids are analyzed, the MALDI-TOF MS typically results in mostly denatured single-strand detection.

The ions generated by the laser pulses are accelerated to a fixed kinetic energy by a strong electric field and then pass through an electric field-free region in vacuum in which the ions travel with a velocity corresponding to their respective mass-to-charge ratios (m/z). The smaller m/z ions will travel through the vacuum region faster than the larger m/z ions thereby causing a separation. At the end of the electric field-free region, the ions collide with a detector that generates a signal as each set of ions of a particular mass-to-charge ratio strikes the detector. Usually for a given assay, 10 to 100 mass spectra resulting from individual laser pulses are summed together to make a single composite mass spectrum with an improved signal-to-noise ratio.

The mass of an ion (such as a charged nucleic acid) is measured by using its velocity to determine the mass-to-charge ratio by time-of-flight analysis. In other words, the mass of the molecule directly correlates with the time it takes to travel from the sample plate to the detector. The entire process takes only microseconds. In an automated apparatus, tens to hundreds of samples can be analyzed per minute. In addition to speed, MALDI-TOF MS has one of the largest mass ranges for mass spectrometric devices. The current mass range for MALDI-TOF MS is from 1 to 1,000,000 Daltons (Da) (measured recently for a protein). [R. W. Nelson et al., "Detection of Human IgM at $m/z \sim 1$ MDa," *Rapid Commun. Mass Spectrom.* 9, 625 (1995)]

The performance of a mass spectrometer is measured by its sensitivity, mass resolution and mass accuracy. Sensitivity is measured by the amount of material needed; it is generally desirable and possible with mass spectrometry to

work with sample amounts in the femtomole and low picomole range. Mass resolution, $m/\Delta m$, is the measure of an instrument's ability to produce separate signals from ions of similar mass. Mass resolution is defined as the mass, m , of a ion signal divided by the full width of the signal, Δm , usually measured between
5 points of half-maximum intensity. Mass accuracy is the measure of error in designating a mass to an ion signal. The mass accuracy is defined as the ratio of the mass assignment error divided by the mass of the ion and can be represented as a percentage.

To be able to detect any point mutation directly by MALDI-TOF mass spectrometry, one would need to resolve and accurately measure the masses of
10 nucleic acids in which a single base change has occurred (in comparison to the wild type nucleic acid). A single base change can be a mass difference of as little as 9 Da. This value represents the difference between the two bases with the closest mass values, A and T (A = 2'-deoxyadenosine-5'-phosphate = 313.19 Da;
15 T = 2'-deoxythymidine-5'-phosphate = 304.20 Da; G = 2'-deoxyguanosine-5'-phosphate = 329.21 Da; and C = 2'-deoxycytidine-5'-phosphate = 289.19 Da). If during the mutation process, a single A changes to T or a single T to A, the mutant nucleic acid containing the base transversion will either decrease or increase by 9 in total mass as compared to the wild type nucleic acid. For mass
20 spectrometry to directly detect these transversions, it must therefore be able to detect a minimum mass change, Δm , of approximately 9 Da.

For example, in order to fully resolve (which may not be necessary) a point-mutated (A to T or T to A) heterozygote 50-base single-stranded DNA fragment having a mass, m , of $\sim 15,000$ Da from its corresponding wild type
25 nucleic acid, the required mass resolution is $m/\Delta m = 15,000/9 \approx 1,700$. However, the mass accuracy needs to be significantly better than 9 Da to increase quality assurance and to prevent ambiguities where the measured mass value is near the half-way point between the two theoretical masses. For an analyte of 15,000 Da, in practice the mass accuracy needs to be $\Delta m \sim \pm 3$ Da = 6 Da. In
30 this case, the absolute mass accuracy required is $(6/15,000)*100 = 0.04\%$. Often a distinguishing level of mass accuracy relative to another known peak in the spectrum is sufficient to resolve ambiguities. For example, if there is a known

mass peak 1000 Da from the mass peak in question, the relative position of the unknown to the known peak may be known with greater accuracy than that provided by an absolute, previous calibration of the mass spectrometer.

5 In order for mass spectrometry to be a useful tool for screening for mutations in nucleic acids, several basic requirements need to be met. First, any nucleic acids to be analyzed must be purified to the extent that minimizes salt ions and other molecular contaminants that reduce the intensity and quality of the mass spectrometric signal to a point where either the signal is undetectable or unreliable, or the mass accuracy and/or resolution is below the value necessary to
10 detect single base change mutations. Second, the size of the nucleic acids to be analyzed must be within the range of the mass spectrometry—where there is the necessary mass resolution and accuracy. Mass accuracy and resolution do significantly degrade as the mass of the analyte increases; currently this is especially significant above approximately 30,000 Da for oligonucleotides (~100
15 bases) Third, because all molecules within a sample are visualized during mass spectrometric analysis (i.e. it is not possible to selectively label and visualize certain molecules and not others as one can with gel electrophoresis methods) it is necessary to partition nucleic acid samples prior to analysis in order to remove unwanted nucleic acid products from the spectrum. Fourth, the mass
20 spectrometric methods for generalized nucleic acid screening must be efficient and cost effective in order to screen a large number of nucleic acid bases in as few steps as possible.

The methods for detecting nucleic acid mutations known in the art do not satisfy these four requirements. For example, prior art methods for mass
25 spectrometric analysis of DNA fragments have focussed on double-stranded DNA fragments which result in complicated mass spectra, making it difficult to resolve mass differences between two complementary strands. See, e.g., Tang et al., Rapid Comm'n. in Mass Spectrometry, 8:183-186 (1994).

Thus, there is a need for cost and time effective methods of detecting
30 genetic mutations using mass spectrometry, preferably MALDI or ES, without having to sequence the genetic material and with mass accuracy of a few parts in 10,000 or better.

SUMMARY OF THE INVENTION

5 The present invention provides methods of and kits for detecting mutations in a target nucleic acid comprising nonrandomly fragmenting said target nucleic acid to form a set of nonrandom length fragments (NLFs), determining masses of members of said set of NLFs using mass spectrometry, wherein said determining does not involve sequencing of said target nucleic acid.

10 In a preferred embodiment, the method of detecting mutations comprises obtaining a set of nonrandom length fragments in single-stranded form. The masses of the members of the set of NLFs can be compared with the known or predicted masses of a set of NLFs derived from a wild type target nucleic acid that is the wild type version of the target nucleic acid that is being screened for mutations. The members of the set of single-stranded NLFs can optionally have one or more nucleotides replaced with mass-modified nucleotides, including mass-modified nucleotide analogs. Another optional aspect of the invention is the
15 inclusion of internal calibrants or internal self-calibrants in the set of nonrandom length fragments to be analyzed by mass spectrometry to provide improved mass accuracy.

The present invention includes a number of nonrandom fragmentation techniques for nonrandomly fragmenting a target nucleic acid.

20 In one embodiment, the nonrandom fragmentation technique comprises hybridizing a single-stranded target nucleic acid to one or more sets of fragmenting probes to form hybrid target nucleic acid/fragmenting probe complexes comprising at least one double-stranded region and at least one single-stranded region, nonrandomly fragmenting said target nucleic acid by cleaving said
25 hybrid target nucleic acid/fragmenting probe complexes at every single-stranded region with at least one single-strand-specific cleaving reagent to form a set of NLFs. The set of fragmenting probes can leave single-stranded regions between double-stranded regions formed by hybridization of said set of fragmenting probes to said target nucleic acid. A single-stranded region comprises a portion of a
30 polynucleotide sequence as small as a single phosphodiester bridge, i.e. the phosphodiester bond across from a nick, to 450 nucleotides in length.

5 The fragmenting probes are oligonucleotides that are complementary to a nucleotide sequence of the target nucleic acid. A set of fragmenting probes can be created such that the nucleotide sequences of the members of the set of fragmenting probes represents the entire complement to the nucleotide sequence of the target nucleic acid. For example, a set of fragmenting probes can provide complete complementary sequence to the target nucleic acid. Alternatively, a set of fragmenting probes, when hybridized to the target nucleic acid, can leave single-stranded regions. Also, one or more sets of fragmenting probes can be used such that the members of one set of fragmenting probes contain nucleotide sequences that overlap with nucleotide sequences of members of a second set of fragmenting probes. In yet another aspect, there are provided two sets of fragmenting probes, where members of the second set of fragmenting probes comprise at least one single-stranded nucleotide sequence complementary to regions of said target nucleic acid that are not complementary to any nucleotide sequences in any members of said first set of fragmenting probes.

15 Once the set(s) of fragmenting probes are hybridized to the target nucleic acid, the single-stranded regions are cleaved using single-strand-specific cleaving reagents, including enzymatic reagents as well as chemical reagents. Single-strand specific chemical cleaving reagents include hydroxylamine, hydrogen peroxide, osmium tetroxide, and potassium permanganate.

20 Yet another nonrandom fragmentation technique comprises providing a single-stranded target nucleic acid, hybridizing the single-stranded target nucleic acid to one or more restriction site probes to form hybridized target nucleic acids comprising double-stranded regions where said restriction site probes have hybridized to said single-stranded target nucleic acid and at least one single-stranded region, nonrandomly fragmenting the hybridized target nucleic acids using one or more restriction endonucleases that cleave at restriction sites within the double-stranded regions. Another variation on this technique involves use of universal restriction probes comprising two regions, the first region being single-stranded and complementary to a specific site within the target nucleic acid, and the second region being double-stranded and containing the restriction recognition site for a particular class IIS restriction endonuclease. Class IIS restriction

endonucleases cleave double-stranded DNA at a specific distance from their recognition site sequence.

Another technique for nonrandom fragmentation comprises fragmenting the target nucleic acid with one or more restriction endonucleases to form a set of NLFs. This and the other forms of nonrandom fragmentation can be combined with direct and indirect capture to a solid support to isolate single-stranded NLFs for mass spectrometric analysis.

Another nonrandom fragmentation technique comprises providing conditions permitting folding of said single-stranded target nucleic acid to form a three-dimensional structure having intramolecular secondary and tertiary interactions, and nonrandomly fragmenting said folded target nucleic acid with at least one structure-specific endonuclease to form a set of single-stranded NLFs. A set of nonrandom length fragments can comprise a nested set of NLFs, wherein each member of the set has a 5' end of the target nucleic acid. The structure-specific endonucleases useful for nonrandom fragmentation comprise any nucleases that cleave at structural transitions within nucleic acids, including: Holliday junctions, single-strand to double-strand transitions, or at the ends of hairpin structures.

Another nonrandom fragmentation method comprises mutation-specific cleavage by hybridizing a target nucleic acid to a set of one or more wild type probes and specifically cleaving at any regions of nucleotide mismatch or base mismatch that form between the target nucleic acid and a wild type probe. The mutation-specific cleavage can be accomplished using a mutation-specific cleaving reagent comprising structure-specific endonuclease or chemical reagents.

The nonrandom fragmentation methods described herein can be combined to form different sets or subsets of nonrandom length fragments. For example, the base mismatch nonrandom fragmentation method using wild type probes can be used in concert with a set of nonrandom length fragments that have already been creating using any one of the other nonrandom fragmentation methods. These nonrandom fragmentation methods can also be combined with isolation methods designed to isolate specific sets of single-stranded nonrandom length fragments, for example, only those NLFs derived from the + strand of the target nucleic

acid. The isolation methods include direct capture of the set of NLFs to a solid support or indirect capture of a set of NLFs to a solid support via a capture probe capable of binding to a solid support via covalent or noncovalent binding. The fragmenting, wild type, restriction site, and universal restriction probes described
5 herein can be also be used as capture probes for isolating a particular set of NLFs.

The isolation methods also comprise the use of a solution of volatile salts to wash away undesired contaminants from the set of NLFs intended for mass determination in the mass spectrometer. The volatile salts are useful for removing background noise and can be easily removed by evaporation of the volatile salts
10 prior to mass spectrometric analysis. Volatile salt solutions can be used in a variety of different methods to prepare organic molecules such as nucleic acids and polypeptides for mass spectrometric analysis. Thus, a method is described herein of decreasing background noise, wherein the method comprises obtaining a sample to be analyzed by a mass spectrometer, washing the sample with a solution of
15 volatile salts, and evaporating the solution of volatile salts from the sample.

The fragmentation and isolation methods separately or together can also be combined with the use of internal self-calibrants to improve the mass accuracy of the mass spectrometric analysis.

The above methods, separately or in combination, can also be combined
20 with the use of mass-modified nucleotides and mass-modified nucleotide analogs incorporated in the target nucleic acid or a set of NLFs to improve mass resolution between mass peaks.

Kits for detecting mutations in one or more target nucleic acids in a sample are also provided. In preferred embodiments, such kits comprise one or more
25 single-stranded target nucleic acids, one or more sets of oligonucleotide probes, wherein each of said probes is complementary to a portion of said single-stranded target nucleic acids, and various cleaving reagents, including single-strand specific cleaving reagents, restriction endonucleases (both Class II and Class IIS), and mutation-specific cleaving reagents. The oligonucleotide probes include
30 fragmenting probes, restriction site probes, and wild type probes. Such kits can also contain a matrix, preferably 3-hydroxypicolinic acid. The kits may also contain volatile salt buffers, and buffers providing conditions suitable for the

enzymatic or chemical reactions described above for nonrandomly fragmenting target nucleic acids and isolating nonrandom length fragments in preparation for mass spectrometric analysis. Additionally, the kits may contain solid supports for purposes of isolating nonrandom length fragments.

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A and 1B display examples of resolved nucleic acid fragments (DNA) in the 20,000 to 30,000 Da range using MALDI-TOF mass spectrometry. Both FIG. 1A and 1B are positive ion mass spectra obtained from 200 fmoles of DNA in 3-HPA (3-hydroxypicolinic acid). Each spectrum is a sum of 100 laser pulses at 266 nm. FIG. 1A: a single-stranded 72-mer which also shows a 71-mer. The FWHM resolution is 240, clearly resolving matrix adducts (labelled M). FIG. 1B: 88-mer parent peak has a resolution of 330.

FIG. 2 is a diagram illustrating the basic steps for mass spectrometric analysis of a nonrandomly-fragmented, double-stranded target nucleic acid.

FIG. 3 is a diagram illustrating the expected mass spectrum for a nonrandomly-fragmented double-stranded target nucleic acid that is a heterozygous mix of wild type and mutant nucleic acid where the mutation is an A to T transversion.

FIG. 4A and 4B illustrate the effect on mass resolution of a mass-substituted base where a T has been replaced by heptyldeoxyuridine during amplification of the mutant region. FIG. 4A depicts a mass spectra of a heterozygous mix of wild type and mutant where A has mutated to T. Spectral peaks are separated by 9 mass units. FIG. 4B depicts a mass spectra of a heterozygous mix of wild type and mutant where A has mutated to T. T has been replaced by heptyldeoxyuridine during amplification of the mutant region. Spectral peaks are now separated by 65 mass units.

FIG. 5 is a diagram illustrating the affect of analyzing only positive strand fragments from a heterozygous sample in reducing the number of total fragments and simplifying the mass spectrum.

FIG. 6 is a diagram illustrating the use of restriction site probes to produce nonrandom fragments from single-stranded target nucleic acid. Note that in the

step of purifying nonrandom length fragments, the small cleaved probes will likely be removed during purification.

5 **FIG. 7A and B** illustrate the use of fragmenting probes in conjunction with single-strand-specific endonuclease to produce nonrandom fragments from single-stranded target nucleic acid.

FIG. 8 is a diagram illustrating the use of fragmenting probes in conjunction with single-strand-specific, base-specific chemical cleavage to produce nonrandom fragments from single-stranded target nucleic acid.

10 **FIG. 9A and B** illustrate the use of fragmenting probes to produce nonrandom fragments from heterozygous, single-stranded target nucleic acid in combination with the use a mismatch-specific cleaving reagent to further fragment the target nucleic acid at the site of a mutation.

15 **FIG. 10** is a diagram illustrating a method of detecting a mutation using mass spectrometric analysis of nonrandomly fragmented mutant and wild-type double-stranded nucleic acids that have been denatured and reannealed and then cleaved at any mismatch regions.

20 **FIG. 11** is a diagram illustrating the effect of analyzing only positive strand fragments from a heterozygous sample in reducing the number of total fragments and simplifying the mass spectrum. In this case the positive strand has been nonrandomly fragmented using both restriction endonuclease treatment and mismatch-specific cleavage.

FIG. 12 is a diagram illustrating the use of structures-specific endonucleases to nonrandomly fragment a folded, single-stranded target nucleic acid.

25 **FIG. 13A and B** illustrate the use of a full length capture probe to isolate and purify a set of single-stranded nonrandom length fragments. Shown in FIG. 13B as an option is a second step involving cleavage at mutation-specific mismatch. This mismatch cleavage is particularly useful for cases where mutant DNA is hybridized to wild type.

30 **FIG. 14** is a mass spectrum of a set of nonrandom length fragments from a target nucleic acid containing a mutation, wherein the target nucleic acid is nonrandomly fragmented with hydroxylamine followed by piperidine, resulting in

16.

mutation-specific cleavage at a mismatch. This mass spectrum illustrates the presence of a nonrandom length fragment of 75 bases in size, that results from mutation-specific cleavage.

5 **FIG. 15** is a mass spectrum illustrating hydroxylamine fragmentation of a wild type control of the mutation-containing target nucleic acid of Fig. 14. This mass spectrum lacks a fragment of 75 bases in size due to the lack of a mutation in the wild type target nucleic acid.

10 **FIG. 16** is a mass spectrum of a mutation-containing target nucleic acid that is specifically cleaved with potassium permanganate at the site of a base mismatch.

FIG. 17 is a mass spectrum of a set of 5 single-stranded nonrandom length fragments from an MNL I digest of a wild type target nucleic acid of 184 nucleotides in length.

15 **FIG. 18** is a magnified mass spectrum of two fragments, both 26 bases in length, identical in nucleotide sequence except for a single G to A point mutation, illustrating clear resolution of the two mass peaks.

DESCRIPTION OF SPECIFIC EMBODIMENTS

20 The present invention, directed to methods of screening target nucleic acids to detect mutations using mass spectrometric techniques to analyze post-amplification nucleic acids, provides the advantages of technical ease, speed, and high sensitivity (minute samples are required). The methods described herein *yield a minimal set of products* with improved mass resolution and accuracy and detailed information about the nature and location of the
25 mutation in the target nucleic acid.

The present invention involves obtaining from a target nucleic acid, using a variety of nonrandom fragmentation techniques, a set of nonrandom length fragments (NLFs) and determining the mass of the members of the set of NLFs.

30 The target nucleic acid can be single-stranded or double-stranded DNA, RNA or hybrids thereof, from any source, preferably from a human source, although any source which one is interested in screening for mutations can be used in the methods described herein. When the target nucleic acid is RNA, the RNA

strand is the + strand. If desired, the target nucleic acid can be an RNA/DNA hybrid, wherein either strand can be designated the + strand and the other, the - strand. The target nucleic acid is generally a nucleic acid which must be screened to determine whether it contains a mutation. The corresponding target nucleic acid derived from a wild type source is referred to as a wild type target nucleic acid. The target nucleic acids can be obtained from a source sample containing nucleic acids and can be produced from the nucleic acid by PCR amplification or other amplification technique. The target nucleic acids are typically too large to analyze directly because current mass spectrometric methods do not have the mass accuracy and resolution necessary to identify a single base change in molecules larger than 100 base pairs. Accordingly, the target nucleic acids must be fragmented.

Nonrandom length fragments are nucleic acids derived by nonrandom fragmentation of a target nucleic acid, and can comprise regions or nucleotide sequences that are single-stranded or double-stranded. Due to the simpler mass spectrum that results from mass analysis of single-stranded nonrandom length fragments, it is preferred to determine the masses of sets of single-stranded nonrandom length fragments. The nonrandom length fragments can also contain mass-modified nucleotides, which can enhance ease of analysis, especially when a point mutation has resulted in a very small mass change (on the order of 9 Da) in a nonrandom length fragment as compared to the corresponding wild type nonrandom length fragment. The methods described herein use mass spectrometry to determine the masses of the set or sets of nonrandom length fragments to detect mutations in a target nucleic acid.

The nonrandom fragmentation techniques of the invention are any methods of fragmenting nucleic acids that provide a defined set of nonrandom length fragments, where that set of nonrandom length fragments may be reproducibly obtained by using the same nonrandom fragmentation method on the same target nucleic acid or its wild type version. The methods used for nonrandom fragmentation are designed to optimize the ease of analyzing the resulting mass spectral data by obtaining a range of fragment sizes that avoids significant overlap of mass peaks. The nonrandom fragmentation techniques of the invention include

18.

digestion with restriction endonucleases, structure-specific endonucleases, and specific chemical cleavage. The enzymatic nonrandom fragmentation techniques include partial digestion with restriction endonucleases or structure-specific endonucleases. Partial cleavage occurs when not every possible cleavage site is

5 cleaved by the cleaving reagents used, whether enzymatic or chemical.

Fragmenting probes used in the invention are nucleic acids comprising a single-stranded nucleotide sequence or region that is complementary to a nucleotide sequence of a target nucleic acid. When fragmenting probes are also used as capture probes (i.e. to bind the fragmenting probe and any complementary

10 nucleic acids hybridized thereto to a solid support), the fragmenting probes comprise a first binding moiety that is capable of binding to a second binding moiety attached to a solid support. Upon hybridization of a set of fragmenting probes and a target nucleic acid, the hybrid can be nonrandomly fragmented using one or more cleaving reagents that specifically cleave single-stranded regions.

15 Restriction site probes are oligonucleotides that when hybridized to single-stranded target nucleic acid at specific complementary sequences form complete double-stranded restriction endonuclease recognition sites cleavable using the restriction endonuclease capable of cleaving at or near the recognition sites formed.

20 Universal restriction probes comprise two regions, the first region being single-stranded and complementary to a specific sequence within the target nucleic acid, and the second region being double-stranded and containing the restriction recognition site for a particular class IIS restriction endonuclease.

Capture probes used in the methods described herein comprise fragmenting

25 probes, restriction site probes, universal restriction probes, and any nucleic acids that are bound to a solid support to isolate sets or subsets of nucleic acids or NLFs. Capture probes can comprise a cleavable linkage or cleavable moiety that can be selectively cleaved to release nucleic acids from a solid support prior to mass spectrometric analysis.

30 Wild type probes are nucleic acids derived from a wild type nucleic acid sequence comprising at least one nucleotide sequence complementary to a nucleotide sequence of a target nucleic acid or a member of a set of NLFs. Wild

type probes can be restriction site probes, fragmenting probes, or capture probes comprising a wild type nucleotide sequence that when hybridized to a complementary mutation-containing region of a target nucleic acid results in a base mismatch bulge or loop structure. Wild type refers to a standard or reference nucleotide sequence to which variations are compared. As defined, any variation from wild type is considered a mutation, including naturally occurring sequence polymorphisms.

The term complementary refers to the formation of sufficient hydrogen bonding between two nucleic acids to stabilize a double-stranded nucleotide sequence formed by hybridization of the two nucleic acids.

A single-stranded region comprises a portion of a nucleotide sequence that is capable of being selectively cleaved by single-strand-specific cleaving reagents or structure-specific endonucleases, wherein the portion of a nucleotide sequence can range in size from a single phosphodiester bridge, i.e. the phosphodiester bond across from a nick, to a nucleotide sequence ranging from one to 450 nucleotides in length which are not hybridized to a complementary nucleotide sequence or region.

The types of mass spectrometry used in the invention include ESI or MALDI, wherein the MALDI method may optionally include time-of-flight. The significant multiple charging of molecules in ESI and the fact that complex mixture analysis is generally required mean that the ESI mass spectra will consist of a great many spectral peaks, possibly overlapping and causing confusion. Because the MALDI MS approach produces mass spectra with many fewer major peaks, this method is preferred.

The methods described herein do not require sequencing of the target nucleic acid (using the sequencing methods that require four different base-specific chain termination reactions to determine the complete nucleotide sequence of a nucleic acid) in order to determine the nature and presence of a mutation within the target nucleic acid.

For an initial mutation screen, a useful range of fragment sizes that will allow detection of a point mutation is around 10 to 100 bases. This size range is where mass spectrometry presently has the necessary level of mass resolution and

accuracy. Thus, the fragmentation methods used in this invention are designed to produce from the target nucleic acid, a set of nonrandom length fragments ranging up to 100 bases in size. For purposes of this invention, fragmentation methods that produce a set of random length fragments are not desirable due to the limited reproducibility of such fragments, the limited information available from mass spectrometry analysis of such fragments, and the likelihood of spectral overlap from randomly generated fragments. For example, nonrandom fragmentation permits determination of the mass, base composition, and location of the set of NLFs relative to the target nucleic acid, whereas random fragmentation methods do not.

Existing mass spectrometric instrumentation in the case of MALDI-TOF MS optimally has a mass accuracy of about 1 part in 10,000 (0.01%), four times what is necessary for detecting a single base change in a 50-base long single-stranded DNA fragment. Utilization of mass-modified nucleotides (to be described later) and nearby masses as internal calibrants, provides optimal resolution and mass accuracy of larger nucleic acids, and can extend the usable mutation detection range up to 100 bases, if not higher. Continued advances in mass spectrometric instrumentation will also push this range higher.

Examples of the resolving capabilities of MALDI-TOF MS are displayed in FIG. 1A and 1B. FIG. 1 shows the positive ion TOF mass spectra obtained from 200 fmoles of DNA in the matrix 3-HPA. FIG. 1A (top figure) shows two single-stranded PCR products of lengths 71 and 72 (mass difference = 305 Da = Adenosine) as well as the 72mer and 72mer + a single matrix adduct (M) (mass difference = 139 Da) to be well resolved (FWHM resolution = 240). FIG. 1B (bottom figure) shows an 88 base length single-stranded product having a resolution of 330. Both spectra display high enough accuracy and resolution to detect a point mutation if one were present.

These unique properties of mass spectrometry, MALDI-TOF MS in particular, to separate nucleic acid fragments and identify their mass exactly and the methods taught herein provide novel methods for the screening of target nucleic acids and identification of changes in base composition that might result from genetic mutation.

IMPROVING MASS ACCURACY BY INTERNAL CALIBRATION AND INTERNAL SELF-CALIBRATION

Mass spectrometers are typically calibrated using analytes of known mass.

5 A mass spectrometer can then analyze an analyte of unknown mass with an associated mass accuracy and precision. However, the calibration, and associated mass accuracy and precision, for a given mass spectrometry system (including MALDI-TOF MS) can be significantly improved if analytes of known mass are contained *within* the sample containing the analyte(s) of unknown mass(es). The
10 inclusion of these known mass analytes within the sample is referred to as use of internal calibrants. External calibrants, i.e. analytes of known mass that are not mixed in with the set of nonrandom length fragments of unknown mass and simultaneously analyzed in a mass spectrometer, are analyzed separately. External calibrants can also be used to improve mass accuracy, but because they are not
15 analyzed simultaneously with the set of fragments of unknown mass, they will not increase mass accuracy as much as internal calibrants do. Another disadvantage of using external calibrants is that it requires an extra sample to be analyzed by the mass spectrometer. For MALDI-TOF MS, generally only two calibrant molecules are needed for complete calibration, although sometimes three or more calibrants
20 are used. All of the embodiments of the invention described herein can be performed with the use of internal calibrants to provide improved mass accuracy.

Using the methods described herein, one can obtain a mass spectrum with numerous mass peaks corresponding to the set of nonrandom length fragments of the gene or target nucleic acid under study. If no mutation is present in the target
25 nucleic acid, all of the mass peaks corresponding to the nonrandom length fragments will be at mass-to-charge ratios associated with the set of NLFs from the wild type target nucleic acid. However, if the target nucleic acid contains a mutation, usually no more than one or two of the mass peaks will be shifted in mass, leaving the majority of mass peaks at unaltered locations. In a preferred
30 embodiment of the invention, a self-calibration algorithm uses these unmutated or nonpolymorphic NLFs for internal calibration to optimize the mass accuracy for analysis of the NLFs containing a mutation, thus requiring no added calibrant(s), simplifying the calibration, and avoiding potential spectral overlaps. In a given

sample, however, it will not be known *a priori* which mass peaks, if any, are altered or shifted from their expected masses for the wild type NLFs.

The self-calibration algorithm begins by dividing up the observed mass peaks into subsets, each subset consisting of all but one or two of the observed mass peaks. Each data subset has a different one or two mass peaks deleted from consideration. For each subset, the algorithm divides the subset further into a first group of two or three masses which are then used to generate a new set of calibration constants, and a second group which will serve as an internal consistency check on those new constants. The internal consistency check begins by calculating the mass difference between the m/z values calculated for the second group of mass peaks and the values corresponding to reasonable choices for the associated wild-type NLFs. The internal consistency check can thus take the form of a chi-square minimization where the key parameter is this mass difference. The algorithm finds which data subset has the lowest sum of the squares of these mass differences resulting in a choice of optimized calibration constants associated with group one of this data subset.

After new self-optimized calibration constants are obtained, the mass-to-charge ratios are determined for the mass peaks omitted from the data subset; these are the nonrandom length fragments suspected to contain a mutation. The differences from the observed mass peaks for the wild type NLFs are then used to determine whether a mutation has occurred, and if so, what the nature of this mutation is (e.g. the exact type of deletion, insertion, or point mutation). This self-calibration procedure should yield a mass accuracy of approximately 1 part in 10,000.

FRAGMENTATION OF TARGET NUCLEIC ACIDS

Fragmentation of a target nucleic acid is important for several reasons. First, fragmentation allows direct analysis of large segments of a gene or other target nucleic acid using a single PCR amplification, eliminating the need to multiplex or run separately many smaller-segment PCR reactions.

Second, sequencing of thousands of bases of a gene or other target nucleic acid, by mass spectrometry or otherwise, is a complex and expensive process.

With current capabilities in MALDI and ESI, it is impractical to sequence nucleic acids greater than 50-100 bases in length. Consequently, in order to rapidly screen large genetic regions or target nucleic acids using mass spectrometric nucleic acid sequencing, an impractical and cumbersome number of independent sequencing reactions are necessary to cover the entire genetic region of interest. Therefore, for screening large genetic regions or target nucleic acids for a wide range of potential mutations using mass spectrometry, fragmentation of amplified target nucleic acids ranging from 100 to 1000 base pairs (bp) facilitates faster screening of larger target nucleic acids or genetic regions of interest.

Sequencing can identify the exact location and nature of a genetic mutation in a target nucleic acid, but requires the use of many primers in many separate reactions. Mutations, especially for heterozygous samples analyzed using fluorescence-based systems, are often difficult to identify with confidence. Using the fragmentation methods described herein, a heterozygous sample would yield two distinct mass spectral peaks, correlating to the different masses of the mutant and wild type nucleic acids. Accordingly, the methods described herein can be used to detect a mutation in a target nucleic acid unambiguously.

Third, mass spectrometric analysis of smaller nucleic acid fragments, ranging in size from 2 to 300 bases, more preferably from 10 to 100 bases in length, is desirable because the smaller nucleic acid fragments result in:

(a) more specific localization of any mutations than for larger sized nucleic acid fragments,

(b) superior mass accuracy and resolution of nucleic acid fragments in this mass range, and

(c) a multiplicity of mass peaks that can be used as internal self-calibration standards, further improving the mass accuracy.

For analysis with MALDI-TOF MS, the goal of fragmentation is to produce a set of nonrandom length fragments ranging in length from 2-300 bases, preferably from 10-100 bases in length. The range of lengths serves to better separate and resolve the fragment peaks in the resulting mass spectrum.

Fragmentation of target nucleic acids larger than 100 bases in length can be accomplished using a number of means, including cleavage with one or more

DNA restriction endonucleases targeting specific sequences within double-stranded DNA, chemical cleavage at structure-specific and/or base-specific locations, polymerase incorporation of modified nucleotides that create cleavage sites when incorporated, and targeted structure-specific and/or sequence-specific nuclease treatment.

An exemplary case is where a larger target nucleic acid, e.g. 500 bases in length, is nonrandomly fragmented to produce 10 to 30 nonrandom length fragments that can all be individually resolved by MALDI-TOF mass spectrometry. Two different nonrandom length fragments having the same number of bases can still be resolved from each other by mass spectrometry when they differ in base composition and consequently in mass. Gel electrophoresis methods typically cannot resolve equivalent length fragments.

For example, for a 5 kilobase pair (kb) target nucleic acid to be fully analyzed, using nonrandom length fragments with an average size of 30 bases, approximately 170 nonrandom length fragments would need to be screened. Typically, the target nucleic acid would be amplified by a number of DNA amplifications, ~10-20, in order to reduce the number of fragments to be analyzed in any given sample. Each amplified target nucleic acid product would be digested using restriction endonucleases, often with four-base recognition sites to produce the optimal size fragments. It is preferable that the fragments vary in size to simplify the mass spectral data, e.g. 32 bp + 28 bp + 27 bp + 37 bp + ..., although, as stated above, nonrandom length fragments of the same size could potentially be analyzed if their base compositions vary enough to minimize spectral overlap.

A schematic of the process along with a hypothetical mass spectrum is shown in FIG. 2. FIG. 2 illustrates a 161 base target nucleic acid that has been PCR amplified and fragmented using restriction endonucleases. The resulting 6 nonrandom length fragments are produced. When the laser desorption process occurs, during MALDI-TOF mass spectrometric analysis, the 6 double-stranded fragments are *mostly* denatured and the resulting 12 single-stranded nonrandom length fragments are ionized and detected. Shown at the bottom of FIG. 2 is a simulated mass spectral data plot with all the mass peaks resolved.

As can be seen in FIG. 2 it is very common that restriction endonuclease treatment will produce a number of complementary fragments with the same number of bases, e.g. two at 19 and two at 32. The presence of these equal-length fragments places higher constraints on the required resolution for distinguishing all of the different peaks. It is also not uncommon for the two equal-length, complementary fragments to have identical or nearly identical mass values, leaving the possibility that two complementary fragments will not be resolvable.

Often samples will be heterozygous, containing a 50% mixture of both the normal wild type nucleic acid and the mutated target nucleic acid. In the case where the target nucleic acid carries a mutation in a heterozygous mix, one would observe a splitting of peaks within the nonrandom length fragments containing the mutation. An example of this splitting is shown in FIG. 3 where an A-T to T-A transversion or base flip has occurred in one copy of the gene. The expected peaks would be half normal height since their concentrations are halved relative to homozygous concentrations. In this case, the difference between mutant and wild type peaks would be ~9 Da which can be resolved in the 32 base long fragment. The presence of wild type peaks provides internal self-calibrants allowing highly accurate mass differences (as opposed to absolute mass) to be used to determine the base composition change.

The methods described herein permit MALDI-TOF MS analysis of nonrandom length fragments which has a mass accuracy of approximately 1 part in 10,000. The use of internal self-calibrants makes it possible to extend this level of accuracy up to and potentially beyond 30,000 Da or 100 bases. This mass accuracy enables exact sizing of nucleic acid fragments and the determination of the presence and nature of any mutation, including point mutations, insertions and deletions, even in a heterozygous environment. Further described herein are methods for improving the resolution of individual fragments by means including elimination of equal-length complementary pairs through the use single-strand-targeted fragmentation and/or isolation procedures, and the incorporation of mass-modified nucleotides to enhance the mass difference between similar sized fragments and/or mutant and wild type fragments. In addition, these methods

provide for the removal of salts and other deleterious materials as well as a means for the removal of unwanted nucleic acid fragments prior to mass spectroscopic analysis.

5 **MASS RESOLUTION, MASS ACCURACY, AND THE USE OF MASS-MODIFIED NUCLEOTIDES**

Any of the embodiments of the invention described herein optionally include nonrandom length fragments having one or more nucleotides replaced with mass-modified nucleotides, wherein said mass-modified nucleotides comprise
10 nucleotides or nucleotide analogs having modifications that change their mass relative to the nucleotides that they replace. The mass-modified nucleotides incorporated into the nonrandom length fragments of the invention must be amenable to the enzymatic and nonenzymatic processes used for the production of nonrandom length fragments. For example, the mass-modified nucleotides must
15 be able to be incorporated by DNA or RNA polymerase during amplification of the target nucleic acid. Moreover, the mass-modified nucleotides must not inhibit the processes used to produce nonrandom length fragments, including, *inter alia*, specific cleavage by restriction endonucleases or structure-specific endonucleases and digestion by single-strand specific endonucleases, whenever such steps are
20 used. Mass-modifications can also be incorporated in the nonrandom length fragments of the invention after the enzymatic steps have been concluded. For example, a number of small chemicals can react to modify specific bases, such as kethoxal or formaldehyde.

Any or all of the nucleotides in the nonrandom length fragments can be
25 mass-modified, if necessary, to increase the spread between their masses. It has been shown that modifications at the C5 position in pyrimidines or the N7 position in purines do not prevent their incorporation into growing nucleic acid chains by DNA or RNA polymerase. [L. Lee et al. "DNA Sequencing with Dye-Labeled Terminators and T7 DNA Polymerase: Effect of Dyes and dNTPs on
30 Incorporation of Dye-Terminators and Probability Analysis of Termination Fragments" Nuc. Acids. Res. 20, 2471 (1992)] For example, an octynyl moiety can be used in place of methyl on thymidine to alter the mass by 94 Da.

Mass-modifying groups can be, for example, halogen, alkyl, ester or polyester, ether or polyether, or of the general type XR, wherein X is a linking group and R is a mass-modifying group. The mass-modifying group can be used to introduce defined mass increments into the nonrandom length fragments. One of skill in the art will recognize that there are numerous possibilities for mass-modifications useful in modifying nucleic acid fragments or oligonucleotides, including those described in *Oligonucleotides and Analogues: A Practical Approach*, Eckstein ed. (Oxford 1991) and in PCT/US94/00193, which are both incorporated herein by reference.

At larger mass ranges (30,000-90,000 Da), the mass resolution and mass accuracy of current MALDI-TOF mass spectrometers will not be sufficient to identify a single base change. For this reason, it may be preferable to increase the useful mass range artificially by substituting standard nucleotides within either a target nucleic acid or a nonrandom length fragment with mass-modified nucleotides having significantly larger mass differentials. Use of mass-modified nucleotides applies as well to the mass range below 30,000 Da. Mass modification can generally increase the quality of the mass spectra by enlarging the mass differences between NLFs of similar size and composition. For example, mass-modified nucleotides can increase the minimum mass difference between two nonrandom length fragments that are identical in base composition except for a single base which is an A in one NLF and is a T in the other. Normally, these two NLFs will differ in mass by only 9 Da. By incorporating a single mass-modified nucleotide into one of the bases, the mass difference can be >20 Da. The spectra in FIG. 4 depict the influence mass-modified nucleotides can have on fragment resolution. One example of the many possible mass modifications useful in this invention is the use of 5-(2-heptynyl)-deoxyuridine in place of thymidine. The replacement of a methyl group by heptynyl changes the mass of this particular nucleotide by 65 Da. An A to T transversion in a nucleic acid fragment in which all thymidine bases have been replaced with 5-(2-heptynyl)-deoxyuridine would produce a peak shift of 56 Da as opposed to 9 Da for the same nucleic acid fragments without the mass-modified nucleotides. The use of mass-modified nucleotides is especially important in the analysis of NLFs derived from RNA. Normally, the masses of C

and U vary by only 1 Da, making it practically impossible to detect C to U or U to C point mutations within a given fragment.

BENEFITS OF ANALYZING SINGLE-STRANDED NUCLEIC ACIDS

5 The goal of this invention is the accurate determination of the masses of a set of resolved nonrandom length fragments and correlation of this data to the characterization of any mutation, if present. The embodiments of this invention include mass spectrometric determination of masses of the members of a set of single-stranded nonrandom length fragments as well as mass determination of the members of a set of mass-modified, double-stranded nonrandom length fragments. 10 The preferred embodiment is to detect mutations in a target nucleic acid comprising obtaining a set of nonrandom length fragments in single-stranded form, wherein the single-stranded nonrandom length fragments are derived from one of either the positive or the negative strand of the target nucleic acid or where the set is a subset of fragments derived from both the positive and the negative strands of 15 the target nucleic acid. The examples of single-stranded methods described herein focus on fragments derived from the positive strand.

 FIG. 2 and 3 illustrate that each double-stranded nonrandom length fragment, comprising two complementary strands, produces two peaks in the mass spectrum corresponding to the denatured single strands. The additional peaks 20 from double-stranded nonrandom length fragments as compared to single-stranded nonrandom length fragments add to congestion of mass peaks in the mass spectra, as well as introducing the possibility that it may be extremely difficult, if not impossible, to resolve the complementary fragments if they have nearly or exactly identical base compositions. Furthermore, some portion of the double-stranded 25 nonrandom length fragments do not fully denature, and mass peaks corresponding to the double-stranded products increase the spectral congestion.

 Because spectra using both strands contain a two-fold redundancy in data, since any mutation in one strand will be present within its complement, it is 30 reasonable to remove one strand prior to mass spectrometric analysis and still produce all of the data necessary for complete mutation analysis. For these reasons, it is the preferred embodiment to analyze a set of single strands where

only one of the two complementary sets nucleic acid fragments representing the full target sequence is used.

FIG. 5 shows the expected spectrum if only the nonrandomly fragmented positive strand of a target nucleic acid from FIG. 3 is analyzed by mass spectrometry. Analysis of one of the two complementary strands of the double-stranded nonrandom length fragments halves the number of expected peaks within the mass spectra, allowing more total fragments to be resolved and the possibility that longer total sized target nucleic acids can be analyzed at one time. Removal of one of the two strands from each nonrandom length fragment eliminates the greatest source of complication for each spectra. A number of methods for isolating and preparing both single-stranded and double-stranded nonrandom length fragments for mass spectrometry are described herein.

METHODS OF NONRANDOM FRAGMENTATION OF TARGET NUCLEIC ACIDS

The methods of the invention all involve obtaining from a target nucleic acid a set of resolvable, nonrandom-length fragments and determining the mass of the members of that set using mass spectrometry without sequencing the target nucleic acid. All of the methods described herein involving mass spectrometry include inter alia two types of mass spectrometry, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF). In addition to the restriction endonuclease approach to nonrandomly fragmenting a target nucleic acid, there are a number of other approaches which are described below.

NONRANDOM FRAGMENTATION USING RESTRICTION SITE PROBES

Target nucleic acid can be nonrandomly fragmented using hybridization to nucleic acid, restriction site probes followed by cleavage with one or more restriction endonucleases the recognition sequences of which are contained in the restriction site probes used. "Restriction site probes" are oligonucleotides that when hybridized to single-stranded target nucleic acid at specific sequences form a complete double-stranded recognition site cleavable using restriction endonucleases. The use of restriction site probes is illustrated in FIG. 6.

The sequence of a wild type target nucleic acid can be analyzed to determine which restriction sites would result in an ideal spread of members of a set of NLFs. The restriction site probes are then made using well-known synthetic techniques. The restriction site probes can range from 6 - 100 nucleotides in length, preferably from 10-30 nucleotides in length. One advantage of using very short restriction site probes is that after cleavage with the selected restriction endonucleases, the mass of the members of the set of NLFs having cleaved restriction site probes attached can be directly determined in the mass spectrometer without requiring an isolating step to remove the cleaved restriction site probes. On the other hand, if the cleaved restriction site probes are intended to be used also as capture probes, then the restriction site probes must either have a first binding moiety that is capable of binding to a second binding moiety attached to a solid support or the restriction site probes must have at least one additional nucleotide sequence that is complementary to another probe that is bound to a solid support. A "capture probe" is an oligonucleotide that comprises a portion capable of hybridizing to a nucleic acid, such as a target nucleic acid or a nonrandom length fragment, and a binding moiety that binds the capture probe to a solid phase, either through covalent binding or affinity binding, or a mixture thereof. A capture probe can itself bind to a solid support via binding moieties (direct capture) or can bind to a solid support via another capture probe that binds to a solid support (indirect capture). Also, when the restriction site probe is also used as a capture probe, the preferred range is from 30-50 nucleotides in length, to stabilize the hybridization of the capture probe. By using larger restriction site probes complementary to singular locations on the target nucleic acid it is possible to prevent a restriction enzyme from cutting at all possible locations in a target nucleic acid where restriction sites for a particular restriction endonuclease appear, e.g. cutting at only 5 or 10 restriction sites within a single-stranded target. This is another tool that can be used to produce the optimal nonrandom length fragment set or subset.

An alternative form of restriction site probe is the universal restriction probe as described by Szybalski. [W. Szybalski "Universal Restriction Endonucleases: Designing Novel Cleavage Specificities by Combining Adapter

Oligodeoxynucleotide and Enzyme Moieties," Gene 40, 169 (1985) (incorporated by reference herein)] These universal restriction probes comprise two regions, the first region being single-stranded and complementary to a specific sequence within the target nucleic acid, and the second region being double-stranded and containing the restriction recognition site for a particular class IIS restriction endonuclease. Class IIS restriction endonucleases cleave double-stranded DNA at a specific distance from their recognition sequence. By using this property, and the universal restriction site probe design, it is possible to nonrandomly fragment a single-stranded DNA target at virtually any sequence, providing the means to better control the selection of fragment sizes. It is also possible to mix standard restriction site probes and universal restriction probes in a single reaction.

In this approach, a positive single-stranded target nucleic acid is hybridized to one or more restriction site probes that are complementary to one or more restriction endonuclease recognition sequences within the target nucleic acid. Upon hybridization of the restriction site probes to the target nucleic acid, hybridized target nucleic acids are formed, comprising double-stranded regions where the restriction site probes have hybridized to the target nucleic acid and at least one single-stranded region where the target nucleic acid remains unhybridized to a restriction site probe. The double-stranded regions of the hybridized target nucleic acids are recognition sites for cleavage by one, two or more restriction endonucleases. After the formation of hybridized target nucleic acids, the hybridized target nucleic acids are digested with one, two or more restriction endonucleases, the recognition sequences of which are contained within the double-stranded regions.

The resulting nonrandom length fragments have at least one cleaved restriction site oligonucleotide probe annealed. In some cases, these cleaved probes will be of a size too small to remain hybridized to the target fragments. These nonrandom length fragments can either be purified with the cleaved restriction site oligonucleotide probes attached, or the NLFs can be purified from the cleaved oligonucleotide restriction site probes. Both types of purification can be accomplished using a variety of techniques known in the art, including filtration, precipitation, or dialysis. The preferred approach is to capture the

NLFs to a solid support. The set of nonrandom length fragments can be directly captured to a solid support themselves using a number of means including a binding moiety such as biotin incorporated at numerous base positions throughout the NLFs. Or the NLFs can be indirectly captured to a solid support via hybridization to one or more capture probes that is itself bound to a solid support. The capture probe can comprise the full-length strand of the target nucleic acid that is complementary to the strand from which the nonrandom length fragments were derived. Alternatively, the capture probes can be a set of capture probes each containing at least one sequence complementary to said nonrandom length fragments.

By combining an asymmetric amplification method to produce single-stranded target nucleic acids with the use of restriction site probes, as described herein, one can produce predominantly the desired set of single-stranded NLFs. The restriction site probes used to produce the recognition sites may copurify with the NLFs but can be designed so that they do not interfere with the majority of the mass spectra. For example, the restriction site probes can be designed so that after cleavage their final sizes are less than 20 bases in length and the nonrandom length fragments can have sizes in the range of 20 to 100 bases.

The methods described above can also be modified with the use of uncleavable restriction probes. These uncleavable probes, synthesized with a restriction endonuclease resistant backbone such as phosphorothioate, boranophosphate, or methyl phosphonate, can be used to keep the target nucleic acid NLFs tethered together following restriction digest and can provide a different approach to purification of the NLFs.

FRAGMENTATION USING FRAGMENTING PROBES AND SINGLE-STRAND-SPECIFIC CLEAVAGE

While the use of restriction endonucleases in various combinations and in multiple digests can be an effective approach to fragmentation of the target nucleic acid, when a target presents long sequence lengths (> 100 bases) that do not contain any restriction sites, alternative nonrandom fragmentation techniques are preferred. Long > 100 base fragments will be difficult to probe with sufficient

mass accuracy to determine if a base change mutation has occurred. One way to control the size of fragments is through the use of fragmenting probes and single-strand-specific endonucleases.

Fragmenting probes are defined as nonrandom length, single-stranded
5 oligonucleotides complementary to selected regions of a single-stranded target nucleic acid, and are used through hybridization to define and differentiate within the target nucleic acid regions that are double-stranded versus regions that remains single-stranded. Following differentiation by hybridization the single-stranded regions are subjected to cleavage. As is the case for all of the methods described
10 here that utilize oligonucleotides, the fragmenting probes may be comprised on DNA, RNA or modified forms of nucleic acid such as phosphorothioates, methyl phosphonates or peptide nucleic acids. Three examples of single-strand-specific nucleases that can be used in these methods are Mung bean nuclease, Nuclease S1, and RNase A. These enzymes cut single-stranded DNA or RNA exclusively and
15 act as both *exo*- and endonucleases.

An example of how these probes and enzymes are used follows. A set of fragmenting probes of defined size and sequence are designed to hybridize to complementary regions of the target nucleic acid. It is preferable that the target
20 nucleic acid be primarily if not entirely single-stranded. Use of a T7 or SP6 RNA polymerase transcription system for final amplification is a simple approach to producing the required single-stranded target nucleic acid. Asymmetric PCR can also be utilized to produce primarily single-stranded target.

FIG. 7 shows how different portions of the single-stranded target nucleic acid are hybridized to the oligonucleotide probes. Following hybridization, any
25 regions of the target nucleic acid that remain single-stranded are cleaved using a single-strand-specific endo/exonuclease, such as S1 Nuclease, Mung bean nuclease, or RNase A. The size of the single-stranded region can be as small as a single phosphodiester bridge, i.e. the phosphodiester bond across from a nick. S1 nuclease is capable of cleaving across from nicks. The end products are double-
30 stranded hybrids comprised of two equal length strands: one strand is a member of the set of nonrandom length fragments derived from the target nucleic acid and the other strand is a member of the set of fragmenting probes, wherein said NLFs are

hybridized to said fragmenting probes. Either these double-stranded hybrids or isolated single-stranded nonrandom length fragments derived from said target nucleic acid can be used for MALDI-TOF mass spectrometric analysis.

Preferably, the analysis of the single-stranded nonrandom length fragments derived from said target nucleic acid provides a simpler mass spectrum. It should be noted that when the complementary strands are a mixed DNA/RNA hybrid there will be a significant mass difference between the two strands in all cases, making each strand more easily resolvable in the mass spectrum.

Unlike the restriction endonuclease nonrandom fragmentation approach, with this method it is possible to use a DNA/RNA hybrid providing a convenient route toward digesting the fragmenting probes after nonrandomly fragmenting the target nucleic acid. Isolation of the set of NLFs from the set of fragmenting probes is another means to simplify the mass spectra. Because of the different chemical nature of the two strands of the hybrid, it is possible to utilize DNA- or RNA-specific enzymes to digest the fragmenting probes. As an example, DNase can be used to digest fragmenting probes comprised of DNA while leaving nonrandom length RNA fragments intact or RNase can be used to digest RNA probes while leaving nonrandom length DNA fragments intact. It is also possible to utilize different chemistries to specifically digest one strand or the other. These chemistries include the use of acid to digest DNA or base to digest RNA as well as a multiplicity of other chemistries that can be used to cut modified versions of DNA or RNA. This differential cutting can be exploited to purify and analyze only one of the two strands as described in a later section.

Thus, another embodiment of this invention is a method of detecting a mutation in a DNA fragment from a DNA/RNA hybrid nucleic acid comprising obtaining a DNA/RNA hybrid wherein the DNA/RNA hybrid comprises a single-strand of a DNA fragment hybridized to a single-strand of a RNA fragment, digesting the single-strand of RNA using a RNA-specific reagent, including RNase or a base, determining the mass of the single-stranded DNA fragment using mass spectrometry, and comparing said mass to a mass of a wild type single-stranded DNA fragment. Another embodiment is a method of detecting a mutation in a RNA fragment from a DNA/RNA hybrid nucleic acid comprising obtaining a

DNA/RNA hybrid wherein the DNA/RNA hybrid comprises a single-strand of a DNA fragment hybridized to a single-strand of a RNA fragment, digesting the single-strand of DNA using a DNA-specific reagent, including DNase or an acid, determining the mass of the single-stranded RNA fragment using mass spectrometry, and comparing said mass to a mass of a wild type single-stranded RNA fragment. These embodiments can also be applied to a set of DNA/RNA hybrids, and using the DNA-specific or RNA-specific digestion to leave a set of nonrandom length fragments consisting of DNA fragments or a set of nonrandom length fragments consisting of RNA fragments.

Complete digestion using restriction endonucleases produces a series of fragments that can be aligned end to end but do not overlap. With the use of fragmenting probes and single-strand-specific cleaving reagents described herein, one can design a set of sequence and size specific fragmenting probes that can be used to produce a set of nonrandom length fragments such that one or more members of the set comprise a nonoverlapping nucleotide sequence and a nucleotide sequence that overlaps with a nucleotide sequence of another member of the set. The example shown in FIG. 7 uses a set of sequence and size specific fragmenting probes that overlap (e.g. split into two sets of hybridization reactions) to produce an overlapping set of nonrandom length fragments. The set of nonrandom length fragments that overlap could be nested. By using a set of overlapping nonrandom length fragments to screen for a mutation, one can more narrowly localize the region containing a mutation. If two overlapping nonrandom length fragments both contain the mutation, as is the case in FIG. 7, it is then known that the mutation exists within the small region of overlap. Conversely, if only one of the overlapping fragments contains a mutation, it is known that the mutation cannot be in an overlapping region. This approach plus the ability to design certain fragmenting probes to be very small in size, e.g. 10 to 20 bases (typical fragmenting probes will be anywhere between 10 and 100 bases in length), allows one to probe genetic regions that are known hot spots for mutation with greater detail.

One variant of this method is to use single-strand-specific chemical reagents as a means for cleaving a target nucleic acid target into a set of nonrandom length

fragments. Several base-specific cleavage chemistries have been identified that cleave the nucleic acid backbone at base-specific sites that are single-stranded and, under optimal conditions, demonstrate zero or extremely reduced cleavage levels at base-specific sites that are double-stranded. As an option the target nucleic acid can be synthesized using one or more modified nucleotides in order to make the backbone more vulnerable to chemical cleavage. By using fragmenting probes to hybridize to a target nucleic acid at all sites except the specific locations where cleavage is desired, it is possible to limit cleavage to these single-stranded sites and create a sequence-specific set of nonrandom length fragments. The method, schematized in FIG. 8, can utilize one of a number of different chemistries that are known to be single-strand specific including hydrogen peroxide cleavage and/or 2-hydroperoxytetrahydrofuran cleavage at C. [P. Richterich et al. "Cytosine specific DNA sequencing with hydrogen peroxide" *Nuc. Acids Res.* 23, 4922 (1995); G. Liang, P. Gannet & B. Gold "The Use of 2-Hydroperoxytetrahydrofuran as a Reagent to Sequence Cytosine and to Probe Non-Watson-Crick DNA Structures" *Nuc. Acids Res.* 23, 713 (1995)]. Target nucleic acids that contain cleavage-modified nucleotides can be made by incorporation of modified nucleotide triphosphates during an amplification or polymerization step.

A second variant of this method is to create heterozygous hybrids between the wild type fragmenting probes and the target nucleic acid. By using fragmenting probes comprised of wild type sequence, any hybrids that form with mutant sequence containing a point mutation will create a base mismatch or bulge. If the mutation is a small insertion or deletion, a looped out sequence will occur. With this heterozygous hybrid, it is possible to use one of the structure-specific enzymes or chemistries described in the following section to create a mutation-specific cleavage at the site of a mutation. An example of the pattern of nonrandom length fragments produced is shown in FIG. 9. This approach permits determination of the type and location of the mutation that has occurred. Also as will be described, performance of a mutation-specific cleavage relaxes the mass accuracy and resolution constraints, thus increasing the useful size range for the nonrandom length fragments to be analyzed with MALDI-TOF mass spectrometry to a range of several hundred bases.

MUTATION-SPECIFIC CLEAVAGE USING STRUCTURE-SPECIFIC ENDONUCLEASES

Another nonrandom fragmentation technique involves the use of mutation-specific cleavage at base mismatch regions, if present, using structure-specific endonucleases or single-strand-specific cleavage. Creation of mismatch regions requires hybridization between a mutation containing, single-stranded target nucleic acid and a set of one or more single-stranded complementary wild type probes derived from wild type sequence. Wild type probes can be restriction site probes, fragmenting probes, or capture probes comprising wild type nucleotide sequence that when hybridized to a complementary mutation-containing region of a target nucleic acid results in a base mismatch bulge or loop structure. A base mismatch will be created at the location of the mutation. In one embodiment, the mutation containing positive strand is hybridized to a complementary wild-type probe that comprises the entire negative strand. In the preferred embodiment, the complex of mutation containing positive strand hybridized to one or more complementary, wild type nucleic acid probes is fragmented using either restriction endonucleases, or fragmenting probes coupled with a single-strand-specific cleavage reagent. Any base mismatch regions between the set of wild type probes and the set of NLFs can be specifically cleaved using one or more mismatch-specific cleaving reagents. Examples of these reagents include: structure-specific endonucleases such as T4 endonuclease VII, RuvC, MutY, or the endonucleolytic activity from the 5'-3' exonuclease subunit of thermostable DNA polymerases, single-strand-specific enzymes such as Mung bean nuclease, S1 nuclease or RNase A, and single-strand-specific chemistries such as hydroxylamine, osmium tetroxide, potassium permanganate, or peroxide modification of unpaired bases followed by a backbone cleaving oxidation step.

This mismatch-specific cleavage is used to cleave the mutation-containing nonrandom length fragment at the site of the mutation, thus producing two smaller fragments from the larger mutation-containing fragment. This approach is an efficient and simple way to identify the exact location of a mutation as well as its type. The mismatch-specific cleavage used in combination with one of the nonrandom fragmentation methods described herein can be used to fragment a

large (> 200 bases), single-stranded target nucleic acid into a set of smaller, mass resolvable nonrandom length fragments.

Like EMC and CCM, the mismatch-specific cleavage approach utilizes a mismatch targeting reagent to cut at the point of mutation. The approach described herein improves upon the gel electrophoresis-based methods by focusing on relatively small fragments that take maximum advantage of the mass spectrometer's ability to detect the exact size of a fragment leading to the identification of the exact location and nature of a mutation. The EMC and CCM methods must be followed by DNA sequencing in order to fully characterize a mutation. Using the methods described herein, a mutation in a target nucleic acid can be detected and its location and nature determined without any sequencing.

An example of how a structure-specific enzyme like T4 endonuclease VII can be used for mismatch-specific cleavage is shown in FIG. 10. The first step involves two amplification reactions. First, a target nucleic acid suspected of containing a mutation is amplified. Second, the corresponding wild type target nucleic acid is amplified to create wild type probes. These two amplification reactions can be performed together in one tube if the target nucleic acid is a heterozygous mixture of mutant and wild type. For certain diagnostic procedures, it may be more efficient to produce the wild type probes separately prior to the screening process. The next steps involve fragmentation of the target nucleic acid, e.g. a multiple digest of the target nucleic acid using more than one restriction endonuclease, and a step in which the fragments are mixed, denatured, and then annealed. The fragmentation and denaturing/annealing steps can occur in either order. The purpose of the denaturing/annealing step is to produce a mixture of hybrid target nucleic acids. In a 50:50 mixture of mutant target and wild type nucleic acids, four different products result: 25% homozygous mutant double-stranded nonrandom length fragments, 25% homozygous wild type double-stranded nonrandom length fragments, and 25% each of the two forms of heterozygous mutant/wild type hybrid nonrandom length fragments. See FIG. 10 (illustrating the use of wild type NLFs as wild type probes to generate a base mismatch with mutant NLFs). The heterozygous nonrandom length fragments contain at least one base mismatch at the site of mutation, i.e. the point(s) of sequence variation

between mutant and wild type. The next step involves treatment of the nonrandom length fragments with a mismatch-specific reagent that cleaves at the site of the base mismatch in the heterozygous mutant/wild type nonrandom length fragments. These new cleavages (the number of cleavage events will depend on the particular enzyme used) typically reduce the nonrandom length fragment containing the mutation into two smaller nonrandom length fragments. The 50% of the mixture that contains the homozygous double-stranded nucleic acid fragments with no mismatches will not be cleaved during the mutation-specific cleavage.

Example schematic mass spectral plots are shown in FIG. 10B. An expected spectrum would show a reduction in the peak size of the nonrandom length fragment containing the base mismatch that is cleaved by the structure-specific endonuclease (e.g. peaks 32+(Mut), 32+(Wt), 32-(Wt), and 32-(Mut)) and the introduction of several smaller peaks at lower masses than the mutant peaks representing the set of heterozygous mutant/wild type NLFs that contain base mismatches (see peaks 8+(Mut), 8+(Wt), 11-, 21-(Wt), 21-(Mut), and 24+). These peaks corresponding to the heterozygous NLFs containing base mismatches are reduced in intensity but continue to be present since only 50% of the molecules exist in the heterozygous form that can undergo the mutation-specific cleavage.

It is possible to bias the population of the different heterozygous/homozygous forms by performing the amplifications of the target nucleic acid asymmetrically. Thus, one can maximize the types of nonrandom length fragments yielding mutational data with the majority of the duplex formed during the annealing process being heterozygous positive (+) strand mutant and negative (-) strand wild type.

While it is possible to observe similar patterns using gel electrophoresis techniques, the mass accuracy obtained by mass spectrometry provides the advantage of accurate determination of the nature of the mutation and the ability to determine the size and order of the two nonrandom length fragments created by the mutation-specific cleavage. In the example in FIG. 10B, the resulting mismatch-specific cleavage fragments are represented by sizes 8, 11, 21, and 24 nucleotides in length. Using electrophoretic techniques, it would be impossible to

differentiate the two mutant forms at 8 and 21 (fragments 24+ and 12- do not possess the mutant base and are identical in heterozygous forms C and D), nor would it be possible to directly determine which fragment is upstream (toward the 5' end) and which fragment is downstream (toward the 3' end), e.g. in the positive strand it is 8+ that is upstream from 24+. By providing exact mass values, mass spectrometry allows these strands to be ordered based on mass value database comparison with the fragments expected from the known sequence of the wild type target nucleic acid. By completely identifying the location and nature of the mutation this mass spectrometric method eliminates any need for sequencing the target nucleic acid.

FIG. 10B shows how the mismatch-specific cleavage event adds complexity to the mass spectra. In the example shown, there are several locations where 2, 3, and even 4 different NLFs have the potential to overlap in the mass spectrum, making the full spectrum difficult to resolve. As discussed previously, and shown in FIG. 5, the mass spectra can be greatly simplified by performing the mass spectrometric analysis on only the + or the - strands of the nonrandom length fragments. For example, FIG. 11 shows the set of nonrandom length fragments that are derived by analyzing only the + positive strand of the mutant target nucleic acid. By eliminating the homozygous nonrandom length fragments that are not mutation-specifically cleaved and removing the negative strand from the mass spectrometric analysis, the total number of nonrandom length fragments to be analyzed can be reduced from 20 to 7, with no two mass peaks having the same number of nucleotides. Of course, in other situations, two peaks may be from nonrandom length fragments of the same length depending on the type of mutation present, but such situations will be infrequent.

This mismatch-specific cleavage, like the incorporation of mass-modified nucleotides, extends the usable mass range of the initial target nucleic acid for mass spectrometric analysis since the primary mass accuracy needs are in determining the reduced mass of the nonrandom length fragments created by the mutation-specific cleavage and not in determining the mass of the other nonrandom length fragments that are unaffected by the mutation-specific cleavage.

It is not always necessary to fragment the target nucleic acid in tandem with mismatch-specific cleavage if the size of the nonrandom length fragments created by the mismatch-specific cleavage is small enough to fall into the usable mass range with the necessary mass resolution and accuracy. Target nucleic acids as large as 200 base pairs will yield at least one nonrandom length fragment created by the mutation-specific cleavage wherein the nonrandom length fragments can be a size less than 100 base pairs, e.g. a 200 bp target nucleic acid with a mutation at position 135 will produce nonrandom length fragments of 65 and 135 after cleavage at the site of base mismatch.

FRAGMENTATION USING STRUCTURE-SPECIFIC ENDONUCLEASES TO CLEAVE A FOLDED TARGET NUCLEIC ACID

Another nonrandom fragmentation method of the invention involves providing a target nucleic acid that is either a positive or a negative single-strand; providing conditions permitting folding of the single-stranded target nucleic acid to form a three-dimensional structure having intramolecular secondary and tertiary interactions, and nonrandomly fragmenting the folded target nucleic acid with at least one structure-specific endonuclease to form a set of single-stranded nonrandom length fragments. A diagram of this procedure is provided in FIG. 12. An example of conditions that permit folding of the single-stranded target nucleic acid are heating to denaturation followed by slow cooling to permit annealing to form a thermodynamically favored secondary and tertiary structure. The structure-specific endonucleases include: T4 endonuclease VII, RuvC, MutY, and the endonucleolytic activity from the 5'-3' exonuclease subunit of thermostable DNA polymerases.

An alternative to the use of structure-specific endonucleases is the use of some of the same single-strand-specific chemical cleavage procedures describe earlier in the text. Because of the higher frequency with which these reagents might cleave relative to the structure-specific endonucleases, it is necessary that the secondary and tertiary structures formed by the single-stranded target be more compact, limiting the access of the chemical reagents to the various reactive nucleotides. Approaches to forming these more compact structures include

performance of the reactions at lower temperature, under higher salt conditions, or the use of RNA versus DNA since RNA is known to form more complete secondary and tertiary structures. Using this method, the cleavage reaction can be run to completion to produce a standard set of nonrandom length fragments or run only partially with the potential of producing a nested set of products that can be analyzed by mass spectrometry or by electrophoresis methods.

PURIFICATION METHODS

When analyzing nucleic acids, including nonrandom length fragments, by mass spectrometry, there are several requirements that need to be met.

First, as has been described earlier, is the need to produce fragments within the resolvable range and high mass accuracy range of the mass spectrometer.

Second, is to eliminate from the sample, nucleic acid fragments that do not contribute to the analysis and may unnecessarily convolute the mass spectra. With analysis methods such as gel electrophoresis, a mixture of specifically labeled nucleic acid fragments (radioactive or by fluorescent tagged) can be visualized in the presence of other unlabeled nucleic acid fragments that comigrate but are invisible and therefore do not convolute analysis of the gel data. The mass spectrometric methods described herein do not use any form of labeling that could render certain fragments invisible, e.g. the negative strand in a double-stranded product, and it is therefore necessary to remove such fragments prior to analysis.

Third, is the need to produce samples of relatively high purity prior to introduction to the mass spectrometer. The presence of impurities, especially salts, greatly affects the resolution, accuracy and intensity of the mass spectrometric signal. Contaminating primers, residual sample genomic DNA, and proteins, all can affect the quality of the mass spectra.

In addition to the three requirements listed above it is also desirable for the methods to be amenable to automation, fast and inexpensive, providing an effective approach for detecting genetic mutations.

Existing purification methods are all designed to work with labeled molecules that were typically analyzed by gel electrophoresis. As well as utilizing labels, electrophoresis is, to a certain degree, tolerant of impurities including salts

and proteins. For mass spectrometric analysis, prior art purification methods such as precipitation combined with vigorous alcohol washes, filtering and dialysis, and ion exchange chromatography are unsatisfactory because they cannot eliminate unwanted nucleic acid fragments and normally do not remove all salts from a sample. Solid phase approaches such as glass bead capture under high salt conditions, biotin/streptavidin binding, direct solid-phase covalent linkage, and capture via hybridization to solid phase bound oligonucleotide probes can be used to eliminate unwanted nucleic acid fragments but typically require high levels of salt during many of the wash steps, rendering the products less pure and compromised for mass spectrometric analysis.

The purifications methods of the present invention are better suited to mass spectrometric analysis of nucleic acids than the prior art methods. First, the methods herein physically isolate selected sets of nucleic acids from a multiplicity of impurities including undesirable nucleic acid fragments, proteins, salts, that would result in a poor quality mass spectrum. Second, the methods optionally use a solution comprising volatile salts such as ammonium bicarbonate, dimethyl ammonium bicarbonate or trimethyl ammonium bicarbonate in any of the steps, including hybridization, endonuclease digestion or washing. These two differences are significant advantages over the prior art because: (1) physical separation of the desired set of nucleic acid fragments for mass spectrometric analysis is better than the labelling methods of the prior art that do not physically separate the target nucleic acids from a variety of other impurities that interfere with an accurate mass spectrum; and (2) the use of volatile salts in any of the steps precludes the need for any wash step known in the prior art to merely remove salts or inorganic ions.

Double Strand Fragment Capture Approaches

There are a number of basic ways to purify DNA restriction products from salts and other small molecules including precipitation, filtering, dialysis, and ion exchange chromatography. While all of these methods are effective, they are not all equally useful for removing amplification primers, residual DNA, i.e. genomic DNA, or any proteins used. In addition, none of the basic approaches meets all of

the requirements of automation, speed and cost. The approach that comes closest is the use of small ion exchange spin columns, which are somewhat expensive and not simple to integrate into an automated setup. These small ion exchange spin columns can, however, produce high quality nucleic acids for mass spectrometric analysis. A better alternative is the use of (magnetic) glass beads to capture/precipitate nucleic acids of a specific size range and allow them to be rigorously washed. However, this method, like all of the other prior art methods described above, does not allow for the removal of unincorporated DNA primer since they are of the same size as the nonrandom length fragments to be analyzed and cannot be simply differentiated.

Another general approach to purification of double-stranded fragments is to directly capture the target nucleic acid and/or a set of nonrandom length fragments by one of three means: (A) hybridization to capture probes comprising a first binding moiety that specifically binds to a second binding moiety attached to a solid phase; (B) binding the target nucleic acid or the members of the set of NLFs each comprising a nucleotide sequence and a first binding moiety to a second binding moiety attached to a solid phase; or (C) direct covalent attachment of the target nucleic acid or the members of the set of NLFs to the solid support. Each of these methods has advantages and disadvantages.

(A) Hybridization to solid support bound capture probes is straightforward, specific, and can be made thermodynamically and kinetically favored by optimizing the size and concentration of the capture probes. Optimization is necessary since the set of NLFs would generally prefer to hybridize to their complements rather than to the capture probes. (This approach also works well for single-strand isolation as described in the following section.) A variation is to bind the probes to the solid phase after hybridization to target. Both biotin/streptavidin and covalent approaches for linking the probes to the solid phase are feasible. The principal concern with this approach is that maintenance of the hybridization, especially during wash steps, requires relatively high level of salts and makes it more difficult to produce a salt-free product for mass spectrometric analysis. Solutions to this problem include the use of relatively long capture probes to increase melting temperatures or the use of volatile salts that can

be removed prior to mass spectrometric analysis. The use of volatile salts is described in more detail elsewhere.

5 (B) Biotin coupling to streptavidin (or avidin) requires that any target nucleic acid or nonrandom length fragment to be captured contain a biotin. It is straightforward to capture the target nucleic acid because biotinylated primers can be used in the PCR amplification. In order to capture all of the fragments after a restriction digest, it is necessary to incorporate biotin into all of the fragments. Three possible routes for biotin labeling are, (1) the inclusion of a biotinylated nucleoside triphosphate during fragment synthesis, (2) the use of a DNA
10 polymerase to fill in at 5' restriction overhangs using a biotinylated nucleoside triphosphate, and (3) the use of ligase to ligate a biotinylated oligonucleotide at the restricted ends of the nonrandom length fragments, where the oligonucleotides are either complementary to the restriction sequence overhangs or are capable of blunt end ligation.

15 Each of the three approaches have their problems but are feasible. Biotins incorporated in method (1) may inhibit the restriction endonucleases to be used and prevent the use of structure-specific nucleases in a second mutation-specific step since the biotin may be recognized as DNA modifications to be excised. Method (2) is more feasible but requires a preliminary cleanup step to exchange the normal
20 triphosphates for biotinylated ones. Restriction sites are limited to enzymes that produce 5' overhangs. Method (3) is more generalizable than (2); its principal weakness is competition with larger fragments that will want to relegate. However, this competition can be overcome by using an excess of the biotinylated linkers.

25 (C) The approach of direct covalent attachment of NLFs or target to a solid support faces many of the same challenges as the biotin/streptavidin approach but also includes the need to design specific, "hot" (i.e. fast and efficient) binding chemistry working with low concentrations of material.

30 The target or members of a set of NLFs can be covalently attached to a solid support using any of the number of methods commonly employed in the art to immobilize an oligonucleotide or polynucleotide on a solid support. The target

or NLFs covalently attached to the solid support should be stable and accessible for base hybridization.

Covalent attachment of the target or NLFs to the solid support may occur by reaction between a reactive site or a binding moiety on the solid support and a reactive site or another binding moiety attached to the target or NLFs or via
5 intervening linkers or spacer molecules, where the two binding moieties can react to form a covalent bond. Coupling of a target or NLF to a solid support may be carried out through a variety of covalent attachment functional groups. Any suitable functional group may be used to attach the target or NLF to the solid
10 support, including disulfide, carbamate, hydrazone, ester, N-functionalized thiourea, functionalized maleimide, streptavidin or avidin/biotin, mercuric-sulfide, gold-sulfide, amide, thiolester, azo, ether and amino.

The solid support may be made from the following materials: cellulose, nitrocellulose, nylon membranes, controlled-pore glass beads, acrylamide gels,
15 polystyrene, activated dextran, agarose, polyethylene, functionalized plastics, glass, silicon, aluminum, steel, iron, copper, nickel and gold. Some solid support materials may require functionalization prior to attachment of an oligonucleotide or capture probe. Solid supports that may require such surface modification include wafers of aluminum, steel, iron, copper, nickel, gold, and silicon. Solid support
20 materials for use in coupling to a capture probe include functionalized supports such as the 1,1'-carbonyldiimidazole activated supports available from Pierce (Rockford, IL) or functionalized supports such as those commercially available from Chiron Corp. (Emeryville, CA). Binding of a target or NLF to a solid support can be carried out by reacting a free amino group of an amino-modified
25 target or NLF with the reactive imidazole carbamate of the solid support. Displacement of the imidazole group results in formation of a stable N-alkyl carbamate linkage between the target or NLFs and the support.

The target or NLFs may also be bound to a solid support comprising a gold surface. The target or NLFs can be modified at their 5'-end with a linker arm
30 terminating in a thiol group, and the modified target or NLFs can be chemisorbed with high affinity onto gold surfaces (Hegner, et al., Surface Sci. 291:39-46 (1993b)).

5 In all of the methods in which a solid-phase approach is used, the double-stranded nonrandom length fragments can be rigorously washed to remove deleterious contaminants. Following washing it is necessary to release these fragments from the solid support for mass spectrometric analysis. The isolation of a set of NLFs may be performed on the same plate that is used within the mass spectrometer. Both the capture probe hybridization and biotin/streptavidin approaches can use heat and/or pH denaturation to disrupt the noncovalent interactions and afford release of the set of NLFs bound to the solid support. Alternatively, a cleavable linkage can be incorporated between the first binding moiety and the NLFs. Any covalent coupling chemistry will need to be either reversible or it will be necessary to include a separate chemically cleavable linkage somewhere within the bound product. It may also be useful to use a chemically cleavable linkage approach with the biotin/streptavidin strategies so that release of the double-stranded fragments can be performed under relatively mild conditions.

10 15 In all cases the cleavable linkage can be located within the linker molecule connecting the biotin and the base (e.g. a disulfide bond in the linker), within the base itself (e.g. a more labile glycosidic linkage), or within the phosphate backbone linkage (e.g. replacement of phosphate with a phosphoramidate).

20 One alternative to these solid-phase approaches described above is to capture the target nucleic acids prior to nonrandom fragmentation with one or more restriction endonucleases. Rigorous washes to remove polymerase, salts, primers and triphosphates required for amplification are followed by treatment with minimal amounts of restriction enzyme under very low salt conditions. This mixture is then directly analyzed in the mass spectrometer. Mass spectrometry can tolerate salts if their concentrations are low enough and a limited class of restriction enzymes can work under very low salt conditions.

25

30 The low salt approach does limit the restriction sites that can be cleaved as part of the methods of detecting mutations. Many restriction endonucleases require a significant level of salt. An attractive alternative to limiting the restriction endonuclease cleavage reactions to low levels of salt is to replace the salts normally used with volatile salts. These salts, such as ammonium bicarbonate, dimethylammonium bicarbonate or trimethylammonium bicarbonate,

can be removed prior to mass spectrometric analysis through simple evaporation. Evaporation can be accelerated by placement of the sample in a vacuum, such as the mass spectrometer sample chamber, or by heating the sample.

5 **APPROACHES TO CAPTURING SINGLE-STRANDED FRAGMENTS**

As described earlier, analysis of single-stranded nonrandom length fragments is generally preferable since it provides a complete set of data with the minimal number of fragments and therefore simplifies the spectra and facilitates an increase in the total length of nucleic acid that can be analyzed in a single assay.

10 A number of approaches, as described above, can be taken toward the production of single-stranded fragments and their purification which includes the elimination of undesired fragments.

15 If DNA restriction endonucleases are used to produce the nonrandom length fragments, it is necessary that the target nucleic acid have a double-stranded form prior to restriction, or more specifically, that the restriction endonuclease recognition sites be located in double-stranded DNA. The alternative to having fully double-stranded DNA prior to restriction is to hybridize restriction site probes to single-stranded DNA, wherein the restriction site probes are complementary to the restriction sites for selected restriction endonucleases.

20 The basic known methods for DNA isolation - precipitation, dialysis, filtration and chromatography do not isolate single-stranded from double-stranded DNA. If these purification methods are employed it is necessary to add a separate step where single-strand isolation is performed.

25 Isolation of a set of single-stranded NLFs can be accomplished using a set of capture probes. "Capture probes" are oligonucleotides or polynucleotides comprising a single-stranded region complementary to at least one nucleotide sequence of the single-stranded NLFs to be isolated and a first binding moiety. The first binding moiety is capable of covalent or noncovalent binding to a second binding moiety attached to a solid support. The capture probes can comprise a set
30 of capture probes, each of which contains single-stranded regions complementary to a corresponding member of a set of NLFs. A capture probe can also comprise a full-length single-stranded target nucleic acid that is complementary to the

nucleotide sequences of the members of a set of NLFs. The capture probes can be bound to a solid support using the methods described above for binding a target or set of NLFs to a solid support.

5 If restriction endonucleases are used to produce nonrandom length fragments from DNA, the preferred method for isolating single-strand fragments from these products is to use a select set of capture probes. In one embodiment the capture probe consists of either full length positive or full length negative strand where the strand has been modified to contain a solid-phase binding moiety. The process using full length negative strand modified to contain a biotin at the 5' end is illustrated in FIG. 13. The capture probe is made and the target nucleic acid is fragmented in two separate reactions. Following inactivation of the restriction enzymes the probe and double-stranded fragments are mixed, denatured and annealed producing a hybrid product of positive strand fragments annealed to full length negative strand capture probe. The capture probe can be bound to the solid phase via a biotin-streptavidin interaction prior to or following of the probe/fragment hybrid. Following the necessary wash steps the fragments are released and analyzed by mass spectrometry. Optionally, the fragments can be probed for a mutation-specific base-base mismatch and fragmented using one of the mismatch specific reagents described earlier. Illustrations of the different spectra produced without and with the optional second step are shown in FIG. 13. Note that after mutation-specific, mismatch-specific cleavage fragments that are distal from the solid phase binding site will be released into solution and washed away, therefore, not analyzed. Loss of these fragments can enhance the ability for mass spectrometry to quickly and easily identify the site of mutation.

25 An alternative approach to using restriction endonucleases is the use of fragmenting probes. These have been described in detail above, and allow the use of a target nucleic acid consisting of either DNA or RNA. The final products, using fragmenting probes and single-strand-specific nucleases, are double-stranded and thus without any additional steps do not themselves produce the set of single-stranded, nonrandom length fragments necessary for analysis. However, there are several approaches that can be used to yield single-stranded nonrandom length fragments.

The first approach for producing single-stranded nonrandom length fragments is useful when the target is RNA and the probes are DNA or visa versa. In this case, the double-stranded products are RNA/DNA hybrids and can be selectively treated with either a DNA or RNA specific nuclease to yield the opposite NLF intact. Acid or base treatments are also an option. These single-stranded products can then be isolated using a number of conventional methods described above.

A second approach to producing single-stranded products for mass spectrometry is to attach the size and sequence specific capture probes to a solid support before or after hybridization to the target nucleic acid and the single-strand-specific cleavage. Since the probes are bound to the solid phase it becomes possible to capture, wash, and then selectively release the nonrandom length target fragments as single-stranded molecules. Following any wash steps, the nonrandom length target fragments are removed from the solid support by denaturation of the double-stranded complex. Once released, the single-stranded fragments can be directly analyzed by the mass spectrometer.

One of skill in the art will know how to use capture probes to capture single-strands of a set of NLFs to a solid support in all the embodiments of this invention. For example biotinylated capture probes can be used to capture single-stranded fragments following cleavage of the target nucleic acid with restriction endonucleases (optionally after neutralizing the restriction endonucleases). The use of capture probes provides a relatively high level of flexibility to select which set of NLFs to analyze at any given time. Large capture probes, capable of hybridizing to all or several different fragments, can be used to capture the fragments correlating to one strand of a target nucleic acid, e.g. a capture probe that is full length negative strand. A short capture probe or combinations of shorter capture probes can be used to selectively choose particular fragments from either strand to analyze in a given mass spectrometric sample. For example, if several fragments share similar sizes it might be preferable to analyze them separately.

As another embodiment, a full length target nucleic acid can be captured before restriction digestion using a capture probe that is nuclease resistant. In this

case it is necessary to modify the capture probe, typically by changing the backbone composition from phosphate to a phosphorothioate, methyl phosphonate or borano-phosphate. [Uhlmann and Peyman, "Antisense Oligonucleotides: A New Therapeutic Principle," Chemical Reviews 90(4):543-584 (1990)

5 (incorporated by reference herein)] These forms of modification limit cutting on the probe strand, resulting only in the nicking of the target molecule to create sequence-specific, nonrandom length fragments without creating any double stranded breaks. By leaving the modified probe strand intact, it is possible to quickly capture the nonrandom length fragments to the solid phase and purify for
10 mass spectrometric analysis.

All of these isolation or purification methods can be utilized in cases where a mutation-specific cleavage event is utilized. In order to present a base mismatch mutation for cleavage, a heterozygous, double-stranded molecule must be present. Typically this means that the fragmenting probe is composed of the wild type
15 sequence and is hybridized to the target nucleic acid fragments containing the potentially mutated target nucleic acid.

VOLATILE SALTS

The methods of this invention include the use of volatile salts, which is an innovative alternative to NaCl, MgCl₂, or other commonly used salts. Volatile
20 salts are any salts that completely evaporate, leaving little or no salt residue in the sample to be analyzed in the mass spectrometer, for example, the isolated set of NLFs. Volatile salts useful in the methods described herein include ammonium bicarbonate, dimethyl ammonium bicarbonate and trimethyl ammonium bicarbonate. These volatile salts are useful in many different aspects of the
25 methods described herein, including use in hybridizing of nucleic acids, washing nucleic acids to remove impurities, and digestion of nucleic acids with endonucleases or other enzymes. Rather than performing washes at reduced levels of nonvolatile salts, which might cause the nonrandom length target fragments to denature from a solid support bound oligonucleotide probe, it is a preferred
30 embodiment to wash support-bound nonrandom length fragments in the presence of relatively high levels of NH₄HCO₃, e.g. 100 mM, and then to evaporate the volatile salt prior to analysis by mass spectrometry. Volatile salts are useful for

buffer exchange in all cases where nucleic acids are to be analyzed by mass spectrometry.

Solid phase purification schemes involving DNA hybridization commonly described in the literature do not focus on the removal of salts since gel electrophoresis techniques are much more tolerant of salts than mass spectrometry. [S. Wang, M. Krinks & M. Moos "DNA Sequencing from Single Phage Plaques using Solid-Phase Magnetic Capture" *Biotechniques* 18, 130 (1995); R.

Sandaltzopoulos & P. Becker "Solid-Phase DNase I Footprinting" *Boehringer Mannheim Biochemica* 4, 25 (1995); both incorporated by reference herein]

These methods are primarily focus on the removal of strands complementary to template prior to enzymatic reaction and/or enzymes and unincorporated labeled nucleotides or primers following reaction. In such schemes residual salt levels can be as high as 100mM NaCl and 25 mM MgCl₂. Mass spectrometry is intolerant of salt concentrations of this level. [T. Shaler et al. "Effect of Impurities on the Matrix-Assisted Laser Desorption Mass Spectra of Single-Stranded Oligodeoxynucleotides" *Anal. Chem.* 68, 576 (1996)] The methods described herein using volatile salts provide an innovative approach to isolating and handling target nucleic acids and/or nonrandom length fragments for mass spectrometric analysis.

The volatile salts can be removed from the sample prior to mass spectrometric analysis by evaporation. Evaporation of the volatile salts can be enhanced using a variety of methods, including use of vacuum, heating, laminar flow of a dry gas over the sample, or, in the case of ammonium bicarbonate (or dimethyl- or trimethylammonium bicarbonate), reduction of the pH by addition of an acid, including 3-HPA, can speed up the decomposition of the salt into ammonia (or dimethyl- or trimethylammonia) and carbon dioxide. Volatile salts can be used in a variety of methods, beyond those described here, for preparing samples of any number of organic molecules, including proteins, polypeptides, and polynucleotides, for mass spectrometric analysis.

Each of the nonrandom fragmentation techniques described herein can be used in combination with any of the isolation methods also described herein. Moreover the nonrandom fragmentation techniques can be used in combination

with each other, as one of ordinary skill in the art using the techniques described herein how to combine the different aspects of the invention. For example, the mutation-specific cleavage technique can be combined with a set of restriction endonuclease-cleaved NLFs. All of these methods and combinations thereof can optionally include use of mass-modified nucleotides, internal calibrants and volatile salts.

The kits described above for nonrandomly fragmenting target nucleic acids and detecting mutations in one or more target nucleic acids can also contain a combination of different means of nonrandomly fragmenting the target nucleic acids as well as different means of isolating the nonrandom length fragments that are to be analyzed by mass spectrometry.

The following examples are provided to illustrate embodiments of the invention, but do not limit the scope of the invention.

EXAMPLES

Example 1. PCR Amplification of Source Nucleic Acids.

PCR methods have been extensively developed during the last decade. An example protocol is as follows. A sample containing 10-10,000 copies of a source DNA molecule is mixed with two antiparallel DNA primers that surround a targeted sequence, e.g. the coding region for a gene involved in carcinogenesis. The PCR mix is composed of: 8 μ l 2.5 mM deoxynucleoside triphosphates, 10 μ l 10X PCR buffer, 10 μ l 25 mM $MgCl_2$, 3 μ l 10 μ M forward primer, 3 μ l 10 μ M reverse primer, 0.3 μ l thermostable Taq DNA polymerase, 64.7 μ l H_2O , and 1 μ l source DNA. The sample tube is sealed and placed into a thermal cycling device. A typical cycling protocol is as follows:

25

Step 1 95°C 2 min.

Step 2 95°C 15 sec.

Step 3 55°C 15 sec.

Step 4 72°C 1 min.

30

Step 5 repeat Steps 2-4 35 times

Step 6 72°C 5 min.

Step 7 stop

Example 2. Production of Single-Stranded Nucleic Acids by Asymmetric PCR.

The basic PCR procedure can be modified in order to produce predominantly one of the two strands. These asymmetric procedures involve modifying the ratios of the two primers, a typical ratio is 10:1.

5

Example 3. Production of Single-Stranded DNA via Biotinylated PCR Products.

For the preparation of capture probes one of the two primers can be synthesized with a biotin moiety internally or at the 5' end of the oligonucleotide. Following a standard PCR, the double-stranded product can be bound to a solid-phase surface coated with streptavidin. For example, 10 pmol of double-stranded PCR product is mixed with 5 μ l MPG [10 mg/ml] paramagnetic streptavidin-coated beads in a binding/washing buffer of 2.0 M NaCl, 10 mM TrisCl, 1 mM EDTA, pH 8.0. The solution is incubated for 15 min. at room temperature with mixing. Following incubation the tube is placed next to a high field, rare earth magnet and the paramagnetic beads with the bound biotinylated PCR product are precipitated to the wall of the tube. The supernatant is removed, and the particles, outside the influence of the magnetic field, are resuspended into binding/washing buffer. The beads and wash solution are mixed and then subjected once again to the magnetic field to precipitate the magnetic particles. The supernatant is once again removed and either the wash step is repeated or the alkaline denaturation step commences. In order to release the unbiotinylated strand from the double-stranded product the beads are mixed with an alkaline denaturation solution, 0.1 M NaOH. The beads are incubated at room temperature for 10 min. which denatures the PCR product and releases the unbiotinylated product into solution. The biotinylated strand, bound to the magnetic beads is precipitated from the solution under the magnetic field and unbiotinylated strand, now single-stranded, is transferred to a new tube with the supernatant. In an optional secondary step, the now single-stranded biotinylated strand can be freed from the magnetic beads by boiling the beads in water for 10 min and transferred with the new supernatant after magnetic precipitation of the magnetic beads.

30

Example 4. Mass Modification of Target Nucleic Acids.

Mass modification of the target nucleic acid is performed during the amplification step. One or more standard deoxynucleoside triphosphates are replaced with modified deoxynucleoside triphosphates. As an example thymidine is replaced with a 5-alkynyl-substituted-2'-deoxyuridine triphosphate. Because the modified nucleotides may not be efficient substrates for DNA polymerase it may be necessary to increase the concentration of the corresponding triphosphate by a factor of 2 to 100 over normal levels.

Example 5. Nonrandom Fragmentation of Double-Stranded Target Nucleic Acids Using Restriction Endonucleases

Specifically-sized, double-strand DNA products produced, for example, by PCR are subjected to sequence-specific fragmentation using restriction endonucleases. As an example, 10 pmoles of a 500 base pair PCR product is treated with one unit each of the frequently cutting enzymes Mnl I and HinP I in the buffer recommended by the enzyme supplier. The reaction is incubated at 37°C for 1 hour, followed by an enzyme-denaturing incubation at 65°C for 15 min.

Example 6. Nonrandom Fragmentation of Single-Stranded Target Nucleic Acids Using Small Oligonucleotide Restriction Site Probes in Combination with Restriction Endonucleases.

Single-stranded DNA target, produced, for example, by asymmetric PCR or by the solid phase methods described in Example 3, is mixed with small oligonucleotide restriction probes complementary to selected restriction site locations. As an example, a set of 10 base long probes targeting the Hae III recognition sequence, are synthesized with the sequence (SEQ ID NO: 1) 5' NNNGGCCNNN 3', where the N's are chosen to allow the restriction site probes to fully complement the single-stranded target DNA at the sites where the Hae III recognition site (e.g. the probe (SEQ ID NO: 2) 5' GACGGCCAAA 3' to complement the target sequence (SEQ ID NO: 3) 5' ...TTTGGCCGTC... 3'). The mixture of target and probes, dissolved in the restriction buffer to be used in the cleavage step, is denatured at 95°C and then

incubated at 32°C (the average T_m melting temperature for the probes) for 15 min. allowing the probes to anneal to target and producing a mixture of single-stranded and double-stranded regions within the target nucleic acid. The hybridized product is then cleaved at the double-stranded sites using one or more specific restriction endonucleases (e.g. Hae III), under conditions similar to those described in Example 3.

Example 7. Nonrandom Fragmentation of Single-Stranded Target Nucleic Acids Using Fragmentation Probes in Combination with Single-Strand-Specific Endonucleases.

Single-stranded DNA target, produced, for example, by asymmetric PCR or by the solid phase methods described in Example 3, are mixed with fragmenting probes complementary to the target DNA. As an example, a mixture of probes with sizes of 24, 26, 28, 30, 32, and 34 each with sequences complementary to different, nonoverlapping regions of the single-stranded target DNA. The mixture of target and probes, dissolved in S1 nuclease digest buffer comprised of 50 mM NaAcetate pH 4.5, 280 mM NaCl, 50 mM $MgCl_2$, and 4.5 mM $ZnSO_4$, are denatured at 95°C and then incubated at 55°C (the average T_m for the probes) for 15 min. allowing the probes to anneal to target and producing a mixture of single-stranded and double-stranded regions within the target nucleic acid. The hybridized product is then digested in the single-stranded regions using 1 U S1 nuclease per μg target DNA, incubated at room temperature for 30 min.

Example 8. Nonrandom Fragmentation of Single-Stranded Target Nucleic Acids Using Mismatch-Specific Cleavage.

Example 8.1. Chemical Cleavage at Mismatched Cytosine

A heterozygous, mutation-containing DNA target is produced, either by PCR of a heterozygous source nucleic acid or by hybridization of wild-type probes to a mutation-containing single-stranded target DNA. For solid phase capture and purification protocols the DNA probes are synthesized either chemically or enzymatically in such a way as to contain biotin moieties. By either route, when a mutation is present a mismatch forms between the target and wild type. A cleavage

5 solution of hydroxylamine is prepared by dissolving 1.39 g of hydroxylamine hydrochloride in 1.6 mL of warm H₂O followed by the dropwise addition of 1.75 mL of diethylamine to yield a solution of pH 6. A 6 mL sample of double-stranded DNA containing a mismatch site is mixed with a 20 mL of hydroxylamine solution and the
10 resulting solution is incubated at 37°C for 30 minutes. The reaction is stopped by the addition of 374 mL of H₂O and the solution is removed either by solid phase capture of the reaction products using magnetic beads with washes performed in a similar manner to that described in Example 3 or by multistep centrifugation in a Microcon-30 ultrafiltration unit (Amicon). The reaction products are redissolved in 45 mL of
15 H₂O and 5 mL of piperidine is added. The solution is incubated at 90°C for 30 minutes and then placed on ice to cool. A 300 mL portion of H₂O is added and samples are either evaporated to dryness or purified by one of the two methods described in Examples 9 and 10.

20 A typical mass spectrum obtained from the hydroxylamine fragmentation at a point mutation is shown in FIG. 14. The source DNA in this case is a section of the coding sequence for the p53 gene. A 134 base long PCR product is produced as in Example 1, amplifying p53 from codon 188 to 233 containing a heterozygous point mutation in codon 213, CGA- > TGA. The forward primer containing a 5'-biotin and a chemically labile linker within the primer, the reverse primer being a standard
25 oligonucleotide. The mismatch containing PCR product is treated with hydroxylamine as described above, cleaving the mismatch at C in codon 213. The product is purified as described in Example 10, and analyzed as described in Example 11. A strong peak appears at the mass correlating to a product 75 bases in size identifying that a C is present in a mismatch in the first position of codon 213. An analysis of mutation-free wild type, shown in FIG. 15, contains no mismatch and therefore no cleavage occurs.

Example 8.2. Chemical Cleavage at Mismatched Thymine

30 DNA is obtained in a similar manner to Example 8.1. The modification reagent is a 20 mM solution of KMnO₄ in deionized H₂O. To 6 mL of double-stranded DNA containing a mismatch site is added 14 mL of the modification reagent. The solution is mixed gently at room temperature over the course of two

minutes during which time the solution turns slightly brown. A 20 mL portion of a solution consisting of 1.25 M sodium acetate pH 8.5 and containing 1 M 2-mercaptoethanol is added to stop the reaction, which results in the solution becoming immediately colorless. A 360 mL portion of H₂O is added and the solution is either spun through a Microcon-30 ultrafiltration unit 2X, collected, and then evaporated to dryness or taken through a solid phase capture and wash protocol. The DNA is redissolved in 45 mL of H₂O and 5mL of piperidine is added. The resulting solution is heated to 90°C for 30 minutes and then placed on ice to cool. After it cools, the solution is diluted by the addition of 300 mL of H₂O and then evaporated to dryness. As an alternative the cleavage products can be purified by one of the two methods described in Examples 9 and 10.

A typical mass spectrum obtained from the KMnO₄ fragmentation at a point mutation is shown in FIG. 16. The source DNA in this case is a section of the coding sequence for the p53 gene. A 134 base long PCR product is produced as in Example 1, amplifying p53 from codon 188 to 233 containing a heterozygous point mutation in codon 213, CGA->TGA. The forward primer containing a 5'-biotin and a chemically labile linker within the primer, the reverse primer being a standard oligonucleotide. The mismatch containing PCR product is treated with KMnO₄ as described above, cleaving the mismatch at C in codon 213. The product is purified as described in Example 10, and analyzed as described in Example 11. A strong peak appears at the mass correlating to a product 75 bases in size identifying that a T is present in a mismatch in the first position of codon 213. Based on the data from the analysis in FIG. 14 and FIG. 16 it is possible to confirm that a C->T mutation has occurred in this p53 sample.

Example 9. Purification of Nonrandom Length Fragments Using Capture Probes

Nonrandom fragments are purified by annealing to a capture probes. The capture probe or probes consists of a sequence or sequences complementary to the selected target nonrandom length fragments. One method uses the a full length capture probe prepared as described in Example 3, another uses a number of chemically synthesized capture probes prepared with biotin covalently attached. For either method the procedure is identical. A 10 µL sample containing a single full-

length biotinylated capture probe or a mixture of smaller, synthetic, biotinylated capture probes is mixed with 10 μ L of nonrandom fragments in an annealing buffer consisting of 300mM NaCl, 10mM Tris, and 1mM EDTA pH 7.5. The mixture is heated in a boiling-H₂O bath for 10 min. and then quickly placed in an ice-H₂O bath.

5 The mixture is then transferred to a pre-heated thermal block at 42 °C (the temperature is adjusted depending on the T_m of the capture probe or probes) and incubated for 1 hour. The solution is then allowed to cool and then mixed with streptavidin-coated magnetic beads. Binding to the beads takes place according to the procedure described in Example 3. After the binding step, in place of the alkaline

10 denaturation step, the bound, hybridized nonrandom fragments are washed with a volatile buffer such as 1 M NH₄HCO₃. After 6 cycles of resuspension in 1 M NH₄HCO₃, magnetic precipitation, and removal of the supernatant, the beads are resuspended in 10 μ L of deionized H₂O and heated to 65°C for 5 min. in order to release the nonrandom fragments from the bound biotinylated strand. The beads are

15 quickly precipitated from the warm solution and the supernatant containing the nonrandom fragments is transferred to another tube. The solution of nonrandom fragments is dried to remove excess volatile buffer and then analyzed by mass spectrometry as described in Example 11.

An example of capture and analysis of nonrandom length fragments is shown

20 in FIG. 17. The source DNA in this case is a section of the coding sequence for the p53 gene. A 184 base long PCR product is produced as in Example 1, amplifying p53 from codon 232 to 292 containing a heterozygous point mutation in codon 248, CGG->CAG. The double-stranded PCR product is digested using the restriction enzyme Mnl I under conditions described in Example 5. A full length capture probe

25 of the negative strand is produced as in Example 3, and the nonrandom length fragments derived from the positive strand are captured and purified as described above. The purified single-stranded fragments are analyzed as described in Example 11. Shown in FIG. 16 are the 5 single-stranded positive fragments produced from an Mnl I digest of the wild type 184 base long PCR product. By performing single-

30 stranded isolation the five similarly sized negative strand fragments are eliminated from the spectra and all of the fragments are fully resolved.

Shown in FIG. 18 is a magnification of the spectra examining the 26 base long fragment that, in the heterozygous mutation case, contains the G->A mismatch. Shown are two clearly resolved peaks with a mass difference of 16 Da, exactly the difference between G and A and thus confirming the presence of a mutation. The third smaller peak correlates to a salt adduct of the high mass 26 base product and emphasizes the need for a process that stringently removes salt prior to analysis.

Example 10. Alternative Purification Method for Mismatch-Specific Nonrandom Length Fragments.

The purification of nonrandom fragments that were produced by a mutation-specific cleavage, e.g. chemical cleavage at mismatch sites, can be achieved in an alternative way. In this case the fragmentation is performed on a PCR product that has one solid- phase capturable strand, e.g. containing biotin, and that is also able to be cleaved from the solid support, e.g. a bridging phosphorothioate linkage contained in the primer region [Mag et al., Nucleic Acids Res. 19(7):1437-1441 (1991)]. As an example of this method, a PCR reaction is performed as described in Example 1, but with one of the primers containing a 5'-end biotin modification and also a bridging phosphorothioate linkage located 3-5 bases from the 3'-end, and the other primer a normal one. After amplification the PCR product is subjected to a mutation-specific fragmentation method directly since, for heterozygous mutations, mismatch-containing heteroduplexes are formed *in situ* during the PCR. In order to check for the possibility of a homozygous mutation, the sample is mixed with an equal amount of wild type control, annealed and then subjected to the fragmentation reaction. The material recovered from the fragmentation reactions is purified and made single-stranded by the method described in Example 3. In this case, after the denaturing step, the products are released from the magnetic beads after several H₂O washes by treatment with 5 μ L of 0.02 mM AgNO₃ and incubating at 45°C for 15 min. The Ag⁺ ions are sequestered by the addition of 1 μ L of 100 mM DTT. The samples are dried to remove excess DTT and then analyzed by mass spectrometry by the method described in Example 11.

Example 11. Mass Spectrometry Analysis.

5 The nucleic acid sample to be analyzed is typically mixed with an equal volume of matrix solution consisting of 0.5 M 3-hydroxypicolinic acid (3-HPA) and 50 mM diammonium hydrogen citrate. Typically a 1 μ L portion of the sample is applied to the mass spectrometer sample stage and allowed to dry under a gentle stream of nitrogen gas at room temperature. When the sample has completely dried to form crystals (typically 5 min.) the sample is inserted into the mass spectrometer for analysis. The usual analysis conditions employ the use of a Nd:YAG laser operating at 266 nm with an average pulse energy of 50 mJ/cm². An average of 100 laser shots is typically used to obtain a spectrum.

10 All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

15 The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the invention and the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: MONFORTE, JOSEPH A.
SHALER, THOMAS A.
TAN, YUPING
BECKER, CHRISTOPHER H.
- 10 (ii) TITLE OF INVENTION: METHODS OF SCREENING
NUCLEIC ACIDS USING MASS
SPECTROMETRY
- (iii) NUMBER OF SEQUENCES: 3
- 15 (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: COOLEY GODWARD LLP
(B) STREET: FIVE PALO ALTO SQUARE
3000 EL CAMINO REAL
(C) CITY: PALO ALTO
(D) STATE: CALIFORNIA
20 (E) COUNTRY: USA
(F) ZIP: 94306
- (v) COMPUTER READABLE FORM:
25 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version
#1.25
- 30 (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: MARCH 4, 1997
(C) CLASSIFICATION:
- 35 (vii) PRIOR APPLICATION DATA
(A) APPLICATION NUMBER: 60/012,752
(B) FILING DATE: MARCH 4, 1996
- 40 (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: JACKIE N. NAKAMURA
(B) REGISTRATION NUMBER: 35,966
(C) REFERENCE/DOCKET NUMBER: GNTR-001/01WO
- 45 (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-843-5214
(B) TELEFAX: 414-857-0663

63.

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA, RNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FEATURE:
(A) NAME/KEY:
(B) LOCATION:

(vi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

NNNGGCCNNN

(3) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA, RNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(v) FEATURE:
(A) NAME/KEY:
(B) LOCATION:

(vi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GACGGCCAAA

(4) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 nucleotides
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA, RNA

- 5 (iii) HYPOTHETICAL: YES
- (iv) ANTI-SENSE: NO
- 5 (v) FEATURE:
- (A) NAME/KEY:
- (B) LOCATION:
- 10 (vi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
- TTTGCCGTC

WE CLAIM:

1. A method of detecting mutations in a target nucleic acid comprising:
obtaining from said target nucleic acid a set of nonrandom length fragments
5 (NLFs) in single-stranded form, wherein said set comprises NLFs
derived from one of either the positive or the negative strand of said
target nucleic acid or said set is a subset of single-stranded NLFs
derived from both the positive and the negative strand of said target
nucleic acid,
10 determining masses of the members of said set using mass spectrometry.
2. The method of claim 1 wherein at least one member of said set of single-
stranded NLFs optionally has one or more nucleotides replaced with mass-modified
nucleotides.
15
3. The method of claim 2 wherein said determining step optionally further
comprises
utilizing internal self-calibrants to provide improved mass accuracy.
- 20 4. The method of claim 3 wherein said target nucleic acid is single-stranded and
said obtaining step further comprises:
hybridizing said single-stranded target nucleic acid to one or more sets of
fragmenting probes to form hybrid target nucleic acid/fragmenting
probe complexes comprising at least one double-stranded region and
25 at least one single-stranded region,
nonrandomly fragmenting said target nucleic acid by cleaving said hybrid
target nucleic acid/fragmenting probe complexes at every single-
stranded region with at least one single-strand-specific cleaving reagent
to form a set of NLFs.
30
5. The method of claim 4 wherein said set of fragmenting probes leaves single-
stranded gaps between double-stranded regions formed by hybridization of said set
of fragmenting probes to said target nucleic acid.

66.

6. The method of claim 5 wherein said hybridizing step further comprises:
providing two sets of single-stranded target nucleic acid and
separately hybridizing a first set of fragmenting probes to a first set of single-
stranded target nucleic acid and a second set of fragmenting probes to
a second set of single-stranded target nucleic acid, wherein said
members of said second set of fragmenting probes comprise at least
one single-stranded nucleotide sequence complementary to regions of
said target nucleic acid that are not complementary to any nucleotide
sequences in any members of said first set of fragmenting probes.

7. The method of claim 6 wherein said members of said first set of fragmenting
probes comprise nucleotide sequences that overlap with nucleotide sequences of said
members of said second set of fragmenting probes.

8. The method of claim 4 wherein said single-strand-specific cleaving reagent is
a single-strand-specific endonuclease.

9. The method of claim 4 wherein said single-strand-specific cleaving reagents
are single-strand specific chemical cleaving reagents.

10. The method of claim 9 wherein said single-strand specific chemical cleaving
reagents are selected from the group consisting of hydroxylamine, hydrogen peroxide,
osmium tetroxide, and potassium permanganate.

11. The method of claim 4 further comprising after said nonrandomly fragmenting
step:

hybridizing one or more of said NLFs to one or more capture probes, wherein
said capture probes comprise a single-stranded region complementary
to at least one of said NLFs and a first binding moiety,

binding said first binding moiety to a second binding moiety attached to a
solid support, wherein said binding occurs either before or after said

67.

hybridizing of said NLFs to one or more capture probes, isolating a set of single-stranded NLFs.

12. The method of claim 4 wherein said fragmenting probes comprise a single-stranded nucleotide sequence and a first binding moiety, further comprising:

after said nonrandomly fragmenting step, binding said first binding moiety to a second binding moiety attached to a solid support, and isolating said set of single-stranded NLFs.

13. The method of claim 3 wherein said obtaining step further comprises: nonrandomly fragmenting said target nucleic acid with one or more restriction endonucleases to form a set of NLFs, hybridizing one or more of said set of NLFs or a subset thereof to one or more oligonucleotide probes, wherein each of said oligonucleotide probes comprises a nucleic acid comprising a single-stranded region and a first binding moiety, binding said first binding moiety to a second binding moiety attached to a solid support either before or after said hybridizing step, and isolating said set or subset of single-stranded NLFs.

14. The method of claim 13 wherein all of said oligonucleotide probes consist of one of either full-length positive or full-length negative single strands of said target nucleic acid and a first binding moiety.

15. The method of claim 13 wherein said binding between said first binding moiety and said second binding moiety is a covalent attachment.

16. The method of claim 13 wherein one binding moiety is a member selected from the group consisting of an antibody, a hormone, an inhibitor, a co-factor portion, a binding ligand, and a polynucleotide sequence, and the other binding moiety is a corresponding member selected from the group consisting of an antigen capable of recognizing said antibody, a receptor capable of recognizing said hormone, an enzyme capable of recognizing said inhibitor, a cofactor enzyme binding site

capable of recognizing said co-factor portion, a substrate capable of recognizing said binding ligand, and a complementary polynucleotide sequence.

17. The method of claim 13 wherein said isolating further comprises:
- 5 washing said set of NLFs bound to said solid support with a solution comprising volatile salts selected from the group consisting of ammonium bicarbonate dimethyl ammonium bicarbonate and trimethyl ammonium bicarbonate.
18. The method of claim 3 wherein said target nucleic acid is single-stranded and wherein said obtaining step further comprises:
- 15 hybridizing said single-stranded target nucleic acid to one or more restriction site probes to form hybridized target nucleic acids having double-stranded regions where said restriction site probes have hybridized to said single-stranded target nucleic acid and at least one single-stranded region, nonrandomly fragmenting said hybridized target nucleic acids using one or more restriction endonucleases that cleave at restriction sites within said double-stranded regions.
19. The method of claim 18 further comprising after said nonrandomly fragmenting step,
- 25 hybridizing said NLFs to one or more capture probes, wherein said capture probes comprise a single-stranded region complementary to at least one of said NLFs and a first binding moiety, binding said first binding moiety to a second binding moiety attached to a solid support, wherein said binding occurs either before or after said hybridizing of said NLFs to one or more capture probes, isolating a set of single-stranded NLFs.
20. The method of claim 19 wherein said cleaved restriction site probes comprise a single-stranded region complementary to half of a restriction endonuclease site and a first binding moiety, and further comprising after said nonrandomly fragmenting
- 30

step, binding said first binding moiety to a second binding moiety attached to a solid support, and isolating a set of single-stranded NLFs.

21. The method of claim 3 wherein said target nucleic acid is single-stranded and said obtaining step further comprises:

5 providing conditions permitting folding of said single-stranded target nucleic acid to form a three-dimensional structure having intramolecular secondary and tertiary interactions,
nonrandomly fragmenting said folded target nucleic acid with at least one
10 structure-specific endonuclease to form a set of single-stranded NLFs,
modifying either said target nucleic acid or said set of single-stranded NLFs such that members of said set of single-stranded NLFs comprise a single-stranded nucleotide sequence and at least one first binding moiety,
binding said first binding moiety to a second binding moiety attached to a solid
15 support, and
isolating said set of single-stranded NLFs.

22. The method of claim 3 wherein said target nucleic acid is single-stranded and said obtaining step further comprises:

20 providing conditions permitting folding of said single-stranded target nucleic acid to form a three-dimensional structure having intramolecular secondary and tertiary interactions,
nonrandomly fragmenting said folded target nucleic acid with at least one
structure-specific endonuclease to form a set of single-stranded NLFs,
25 hybridizing one or more of said set of NLFs to one or more capture probes, wherein said capture probes comprise a single-stranded nucleotide sequence and a first binding moiety,
binding said first binding moiety to a second binding moiety attached to a solid support either before or after said hybridizing step, and
30 isolating a set of single-stranded NLFs.

23. The method of claim 21 wherein said isolated set of single-stranded NLFs comprise any NLFs having a 5' end of said target nucleic acid.

5 24. The method of claim 22 wherein said isolated set of single-stranded NLFs comprise any NLFs having a 5' end of said target nucleic acid.

25. The method of claim 21 wherein said structure-specific endonuclease is selected from the group consisting of:
10 T4 endonuclease VII, RuvC, MutY, and the endonucleolytic activity from the 5'-3' exonuclease subunit of thermo-stable polymerases.

26. The method of claim 3 wherein said target nucleic acid is single-stranded and wherein said obtaining step further comprises:
15 hybridizing said single-stranded target nucleic acid to one or more wild type probes,
nonrandomly fragmenting said target nucleic acid with one or more mutation-specific cleaving reagents that specifically cleave at any regions of nucleotide mismatch that form between said target nucleic acid and any
20 of said wild type probes.

27. The method of claim 26 wherein said nonrandomly fragmenting step further comprises:
digesting said first set of nonrandom length fragments with one or more
restriction endonucleases or
25 cleaving said first set of nonrandom length fragments with one or more single-strand-specific cleaving reagents.

28. The method of claim 26 wherein members of said set of single-stranded NLFs comprise a single-stranded region and at least one first binding moiety, further comprising after said nonrandomly fragmenting step, binding said first binding moiety to a second binding moiety attached to a solid support, and isolating a set of single-stranded NLFs.

29. The method of claim 26 wherein said obtaining step further comprises:
hybridizing members of said set of NLFs to one or more capture probes,
wherein said capture probes comprise a single-stranded nucleotide
sequence and at least one first binding moiety, binding said first
binding moiety to a second binding moiety attached to a solid support,
and isolating a set of single-stranded NLFs.

30. The method of claim 26 wherein said obtaining step further comprises:
isolating a set of single-stranded NLFs comprising any NLFs having a 5' end
of said target nucleic acid.

31. A method of detecting mutations in a target nucleic acid comprising:
nonrandomly fragmenting said target nucleic acid with one or more restriction
endonucleases to form a set of double-stranded NLFs, wherein said
nonrandomly fragmenting further comprises using volatile salts in a
restriction buffer, determining masses of the members of the set of
double-stranded NLFs, wherein said determining does not involve
sequencing of said target nucleic acid.

32. A method of detecting mutations in a double-stranded target nucleic acid comprising:

nonrandomly fragmenting said target nucleic acid using one or more
restriction endonucleases to form a first set of nonrandom length
fragments (NLFs),

hybridizing members of said first set of NLFs to a set of wild type
probes,

72.

nonrandomly fragmenting one or more members of said set of NLFs
with one or more mutation-specific cleaving reagents that
specifically cleave at any regions of nucleotide
mismatch that form between members of said first set
of NLFs and complementary members of said set of
wild type probes, wherein said nonrandomly
fragmenting step forms a second set of NLFs, and
determining masses of members of said second set of NLFs using mass
spectrometry, wherein said determining does not require sequencing of
said target nucleic acid.

33. The method of claim 32 further comprising
obtaining said set of wild type probes by nonrandomly fragmenting a wild
type target nucleic acid using the same restriction endonucleases used
to form said first set of NLFs.

34. The method of claim 33 wherein said steps of nonrandomly fragmenting of
said target nucleic acid and obtaining said set of wild type fragmenting probes are
performed simultaneously in a single solution.

35. The method of claim 32 further comprising before said determining step,
isolating said second set of NLFs wherein said members of said second set
comprise double-stranded nucleotide sequences and a first binding
moiety, and binding said first binding moiety to a second binding
moiety attached to a solid support.

36. The method of claim 32 further comprising before said determining step,
isolating said second set of NLFs wherein said isolating comprises hybridizing
members of said second set of NLFs to one or more capture probes,
wherein said capture probes comprise a single-stranded nucleotide
sequence and a first binding moiety, binding said first binding moiety
to a second binding moiety attached to a solid support.

37. A method of detecting mutations in a target nucleic acid comprising:
nonrandomly fragmenting said target nucleic acid, using a solution comprising
one or more volatile salts to form a set of nonrandom length fragments
(NLFs),
5 determining masses of members of said set of NLFs using mass spectrometry,
wherein said determining does not involve sequencing of said target
nucleic acid.
38. A method of decreasing background noise comprising
10 obtaining a sample to be analyzed by a mass spectrometer,
washing said sample with a solution of volatile salts, and
evaporating the solution of volatile salts from the sample.
39. A method of obtaining nonrandom length fragments from a target nucleic acid
15 comprising:
hybridizing one or more sets of fragmenting probes to said target nucleic acid
to form a set of hybrids,
cleaving single-stranded regions of members of said set of hybrids.
40. A kit for detecting mutations in one or more target nucleic acids in a sample
20 comprising:
(a) one or more sets of fragmenting probes, wherein said fragmenting
probes are complementary to a sequence of one or more of said target
nucleic acids;
25 (b) a single-strand specific cleaving reagent; and
(c) a solid support capable of isolating said single-stranded target nucleic
acids that have been nonrandomly fragmented into single-stranded
nonrandom length fragments.
41. The kit of claim 40, wherein said single-strand specific cleaving reagent is a
30 single-strand-specific chemical cleaving reagent selected from the group consisting of
hydroxylamine, hydrogen peroxide, osmium tetroxide, and potassium permanganate.

74.

42. The kit of claim 40, wherein said single-strand specific cleaving reagent is a nuclease selected from the group consisting of Mung bean nuclease, Nuclease S1, and RNase A.

5 43. A kit for detecting mutations in one or more target nucleic acids in a sample comprising:

- (a) one or more sets of restriction site probes, wherein said probes comprise a single-stranded sequence capable of hybridizing to a sequence of said one or more target nucleic acids;
- 10 (b) one or more restriction endonucleases that cleave at restriction sites within said restriction site probes; and
- (c) a solid support capable of isolating said single-stranded target nucleic acids that have been nonrandomly fragmented into single-stranded nonrandom length fragments.

15

44. The kit of claim 43, wherein said restriction endonuclease is a Class IIS restriction endonuclease.

20

45. The kit of claim 43, wherein said restriction site probe comprises two regions, a first region that is single-stranded and complementary to a specific sequence within said target nucleic acid, and a second region that is double-stranded and contains a restriction recognition site for a Class IIS restriction endonuclease.

1/22

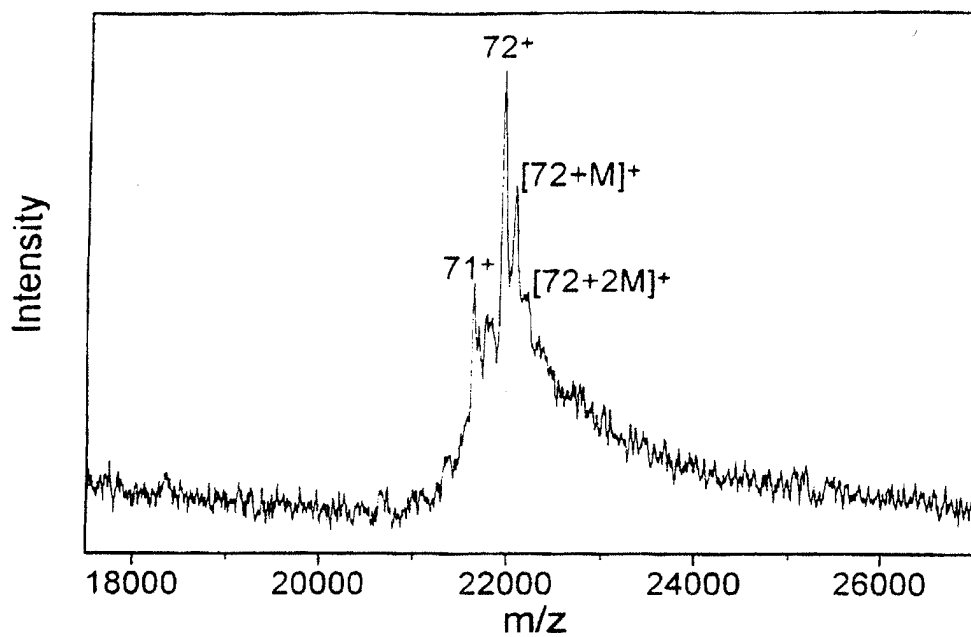


Fig. 1A

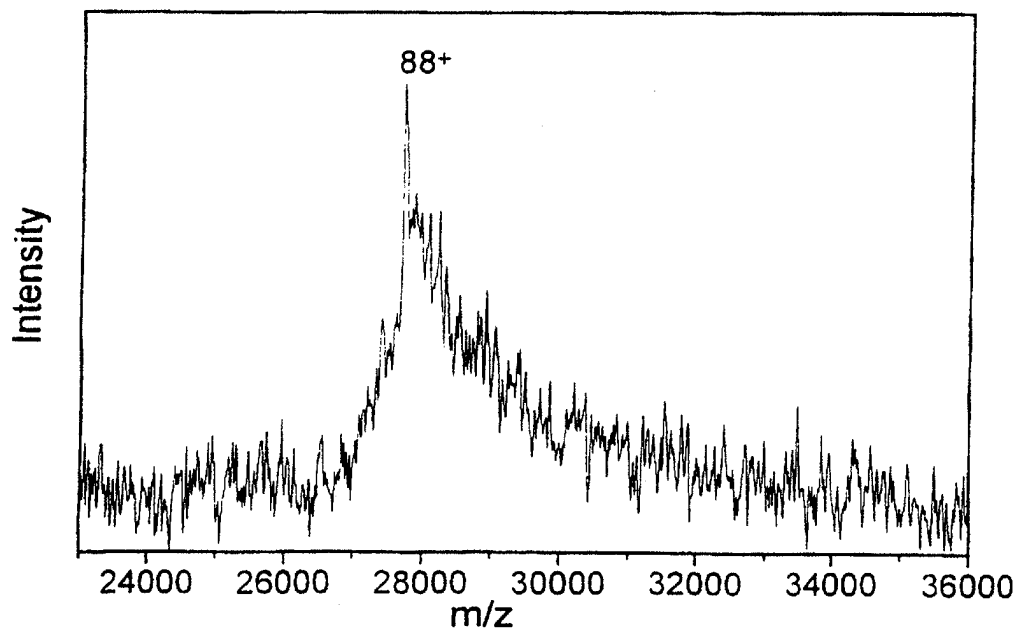


Fig. 1B

SUBSTITUTE SHEET (RULE 26)

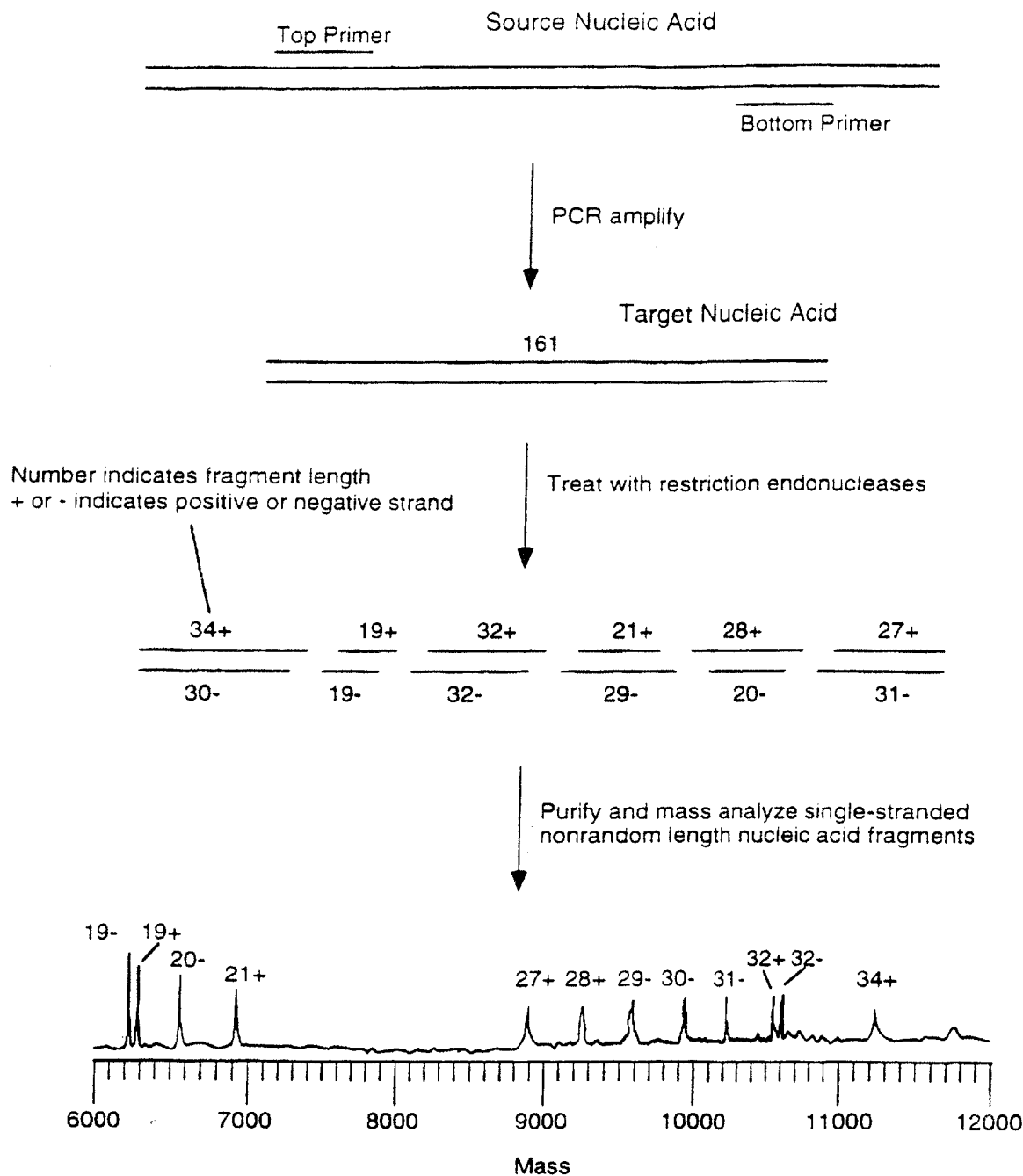
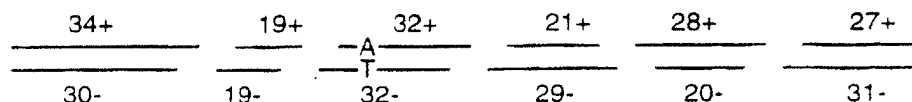


Fig. 2

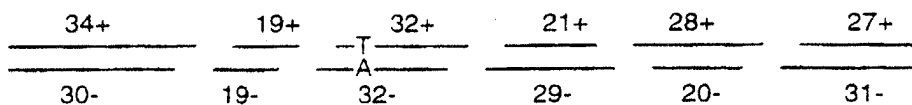
SUBSTITUTE SHEET (RULE 26)

Heterozygous mix

Wild type



Mutant (A to T transversion)



+

Purify and mass analyze single-stranded
nonrandom length nucleic acid fragments

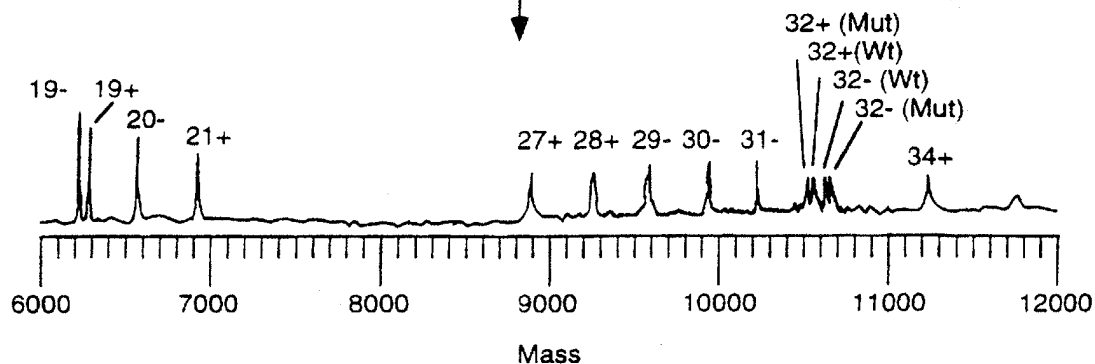


Fig. 3

SUBSTITUTE SHEET (RULE 26)

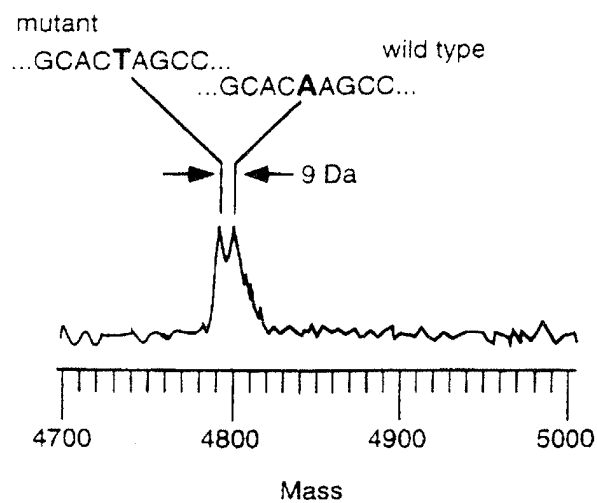


Fig. 4A

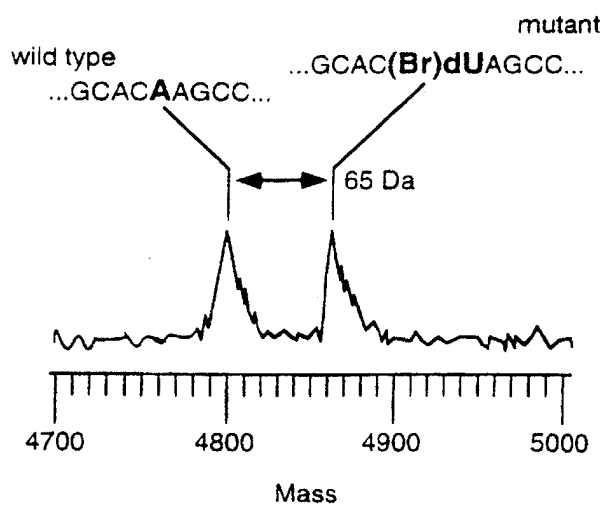


Fig. 4B

SUBSTITUTE SHEET (RULE 26)

Heterozygous mix (positive strand only)

Wild type

34+ 19+ -A- 32+ 21+ 28+ 27+

Mutant (A to T transversion)

34+ 19+ -T- 32+ 21+ 28+ 27+

Purify and mass analyze single-stranded
nonrandom length nucleic acid fragments

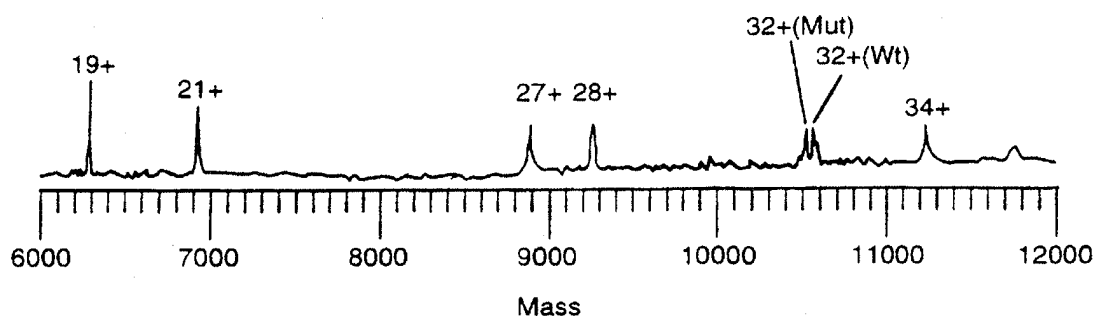
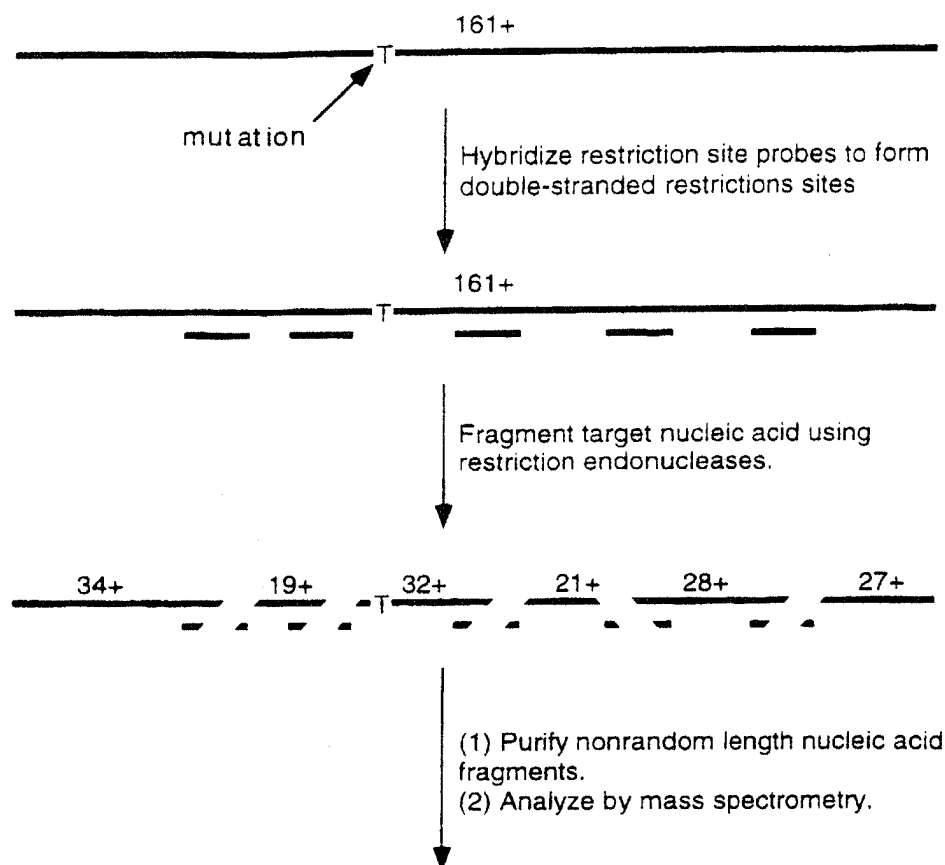


Fig. 5

SUBSTITUTE SHEET (RULE 26)

6/22

Single-stranded nucleic acid target.



Standard expected spectra

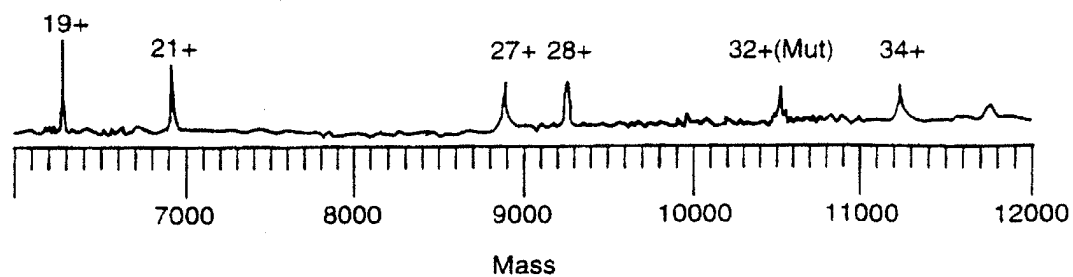


Fig. 6

SUBSTITUTE SHEET (RULE 26)

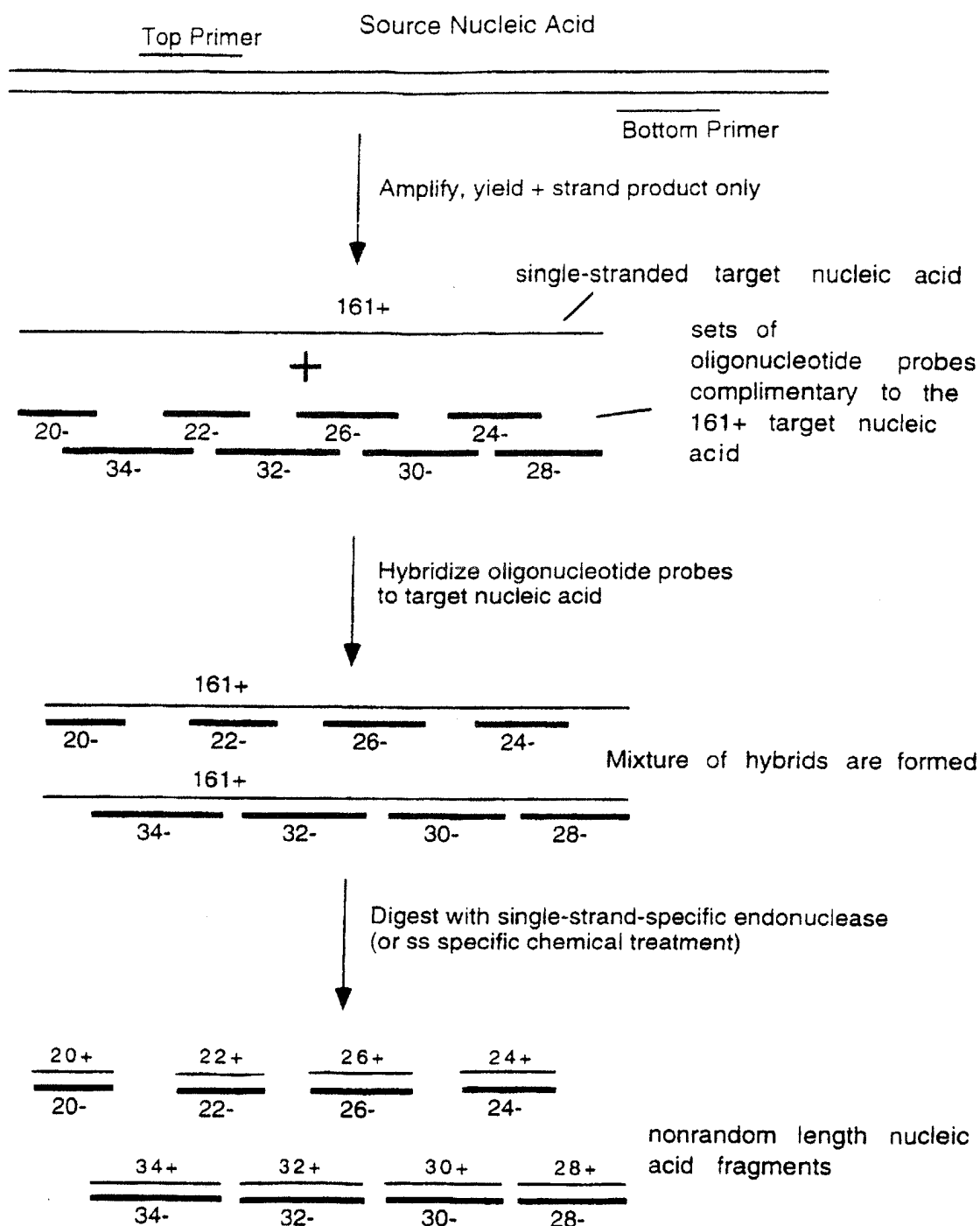


Fig. 7A

SUBSTITUTE SHEET (RULE 26)

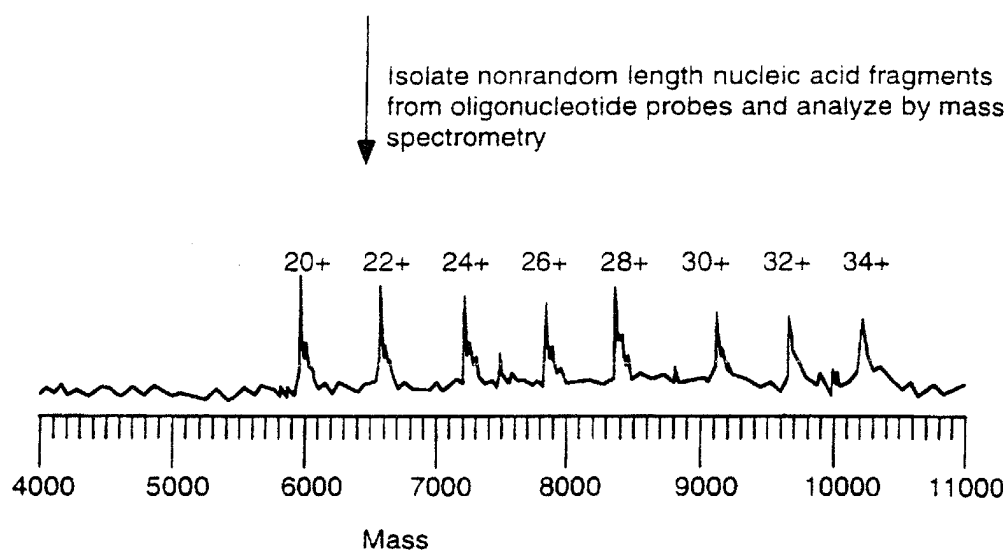


Fig. 7B

9/22

Single-stranded target nucleic acid

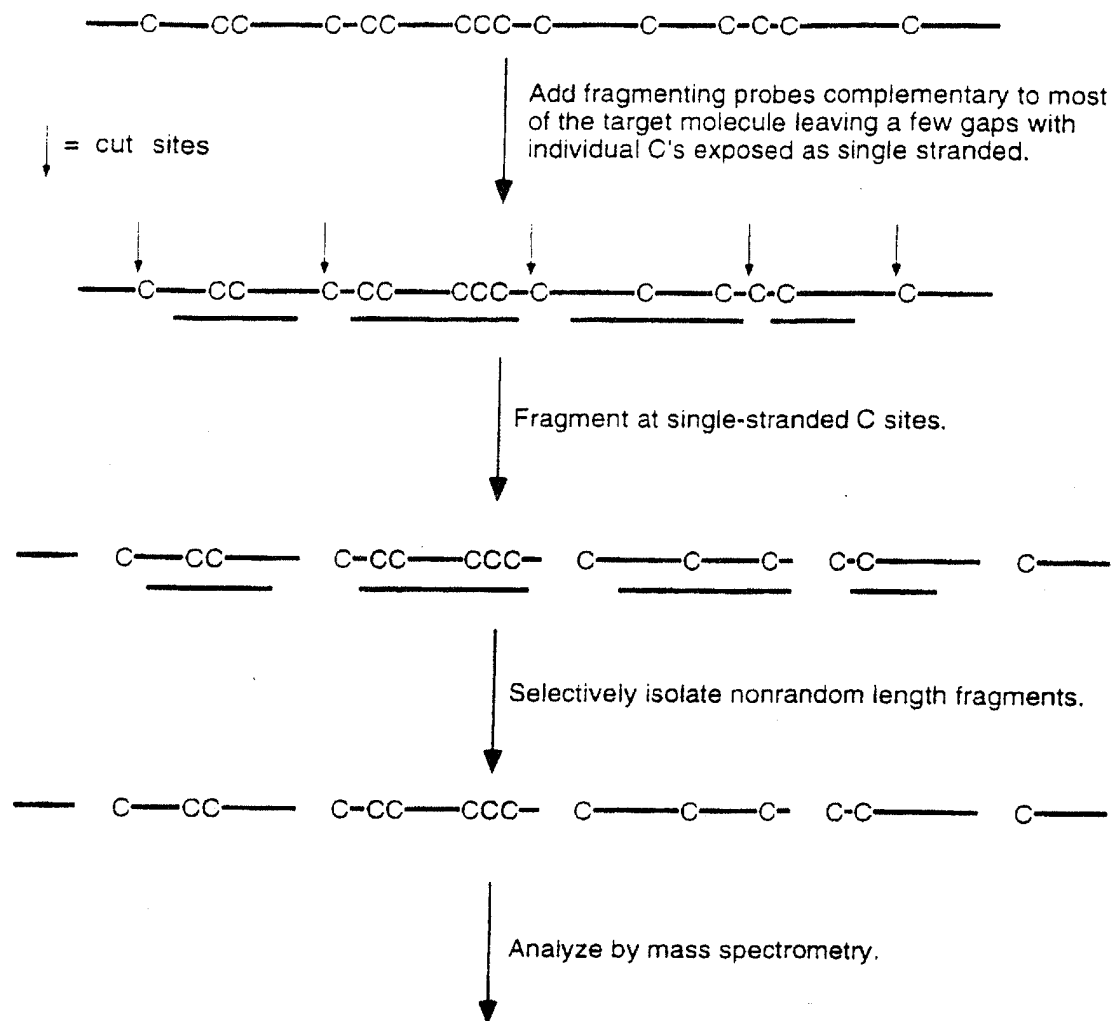


Fig. 8

SUBSTITUTE SHEET (RULE 26)

10/22

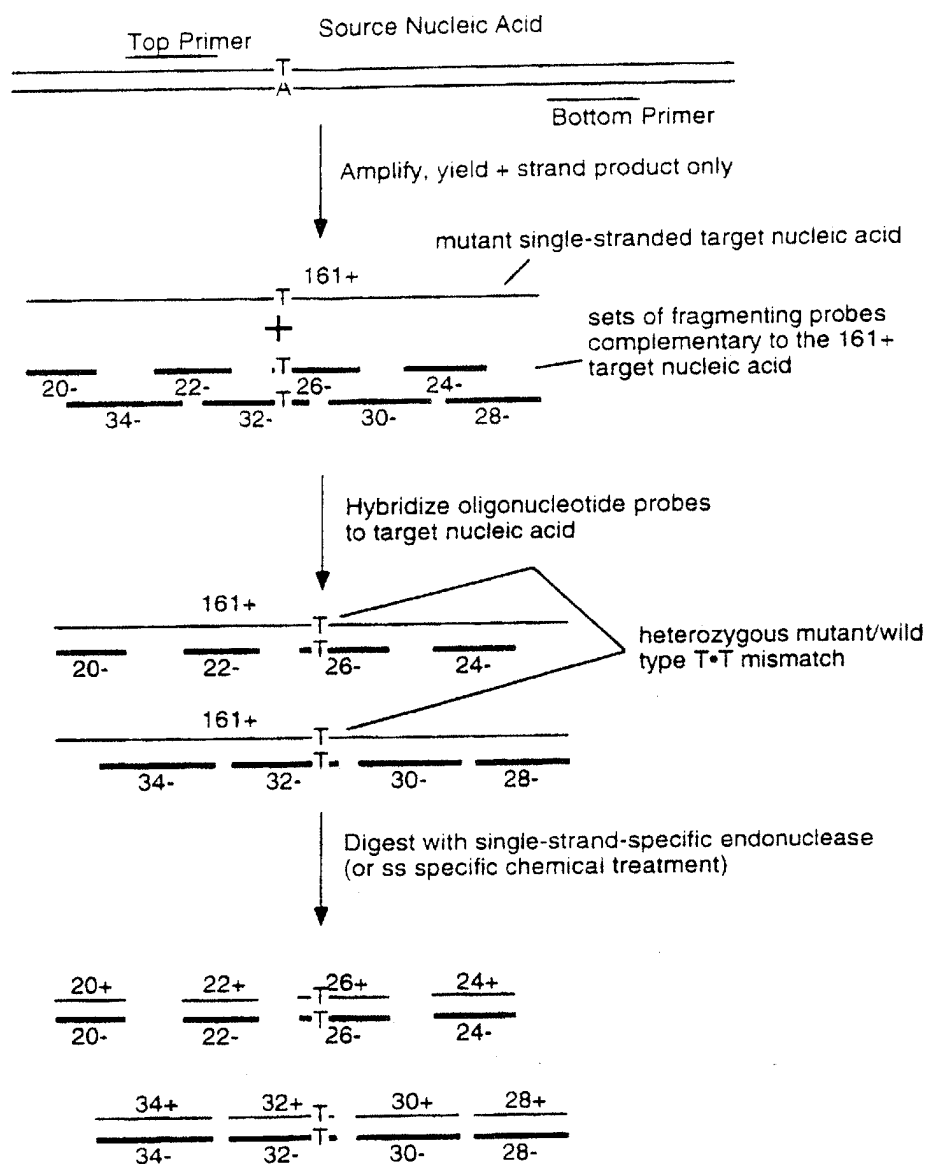


Fig. 9A

SUBSTITUTE SHEET (RULE 26)

11/22

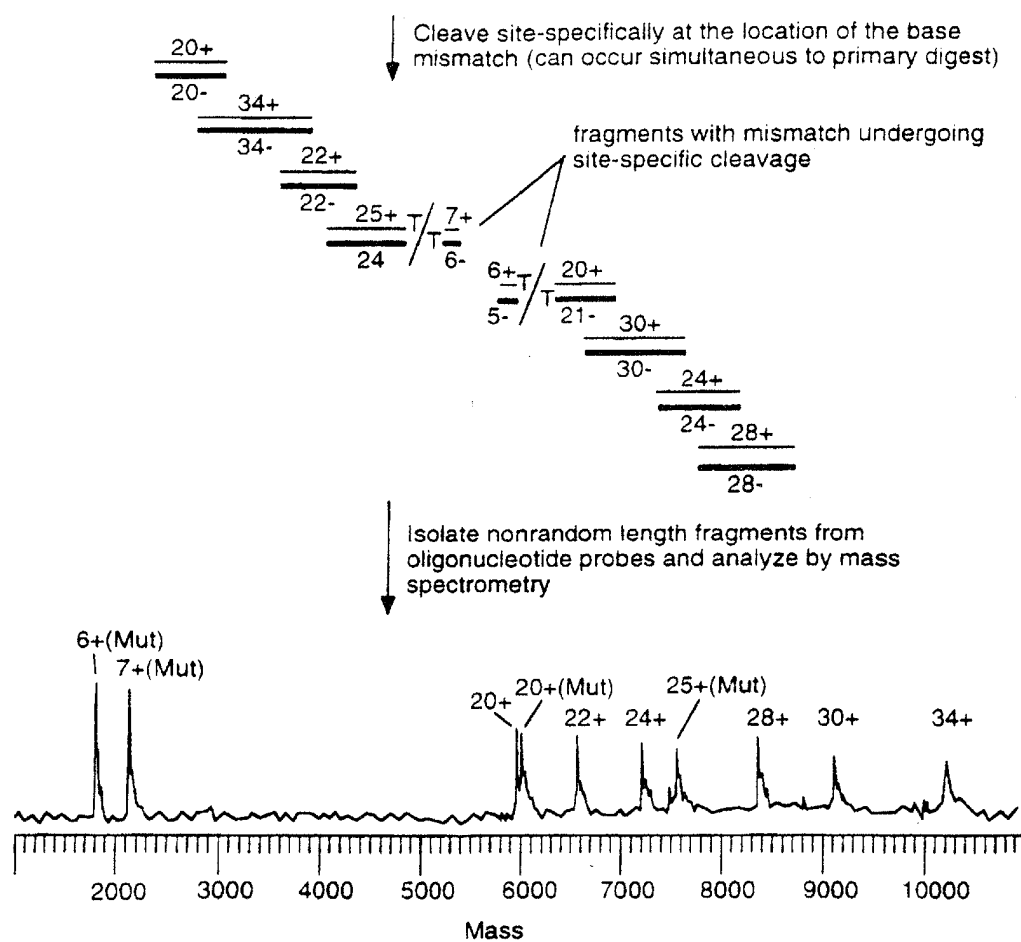


Fig 9B

SUBSTITUTE SHEET (RULE 26)

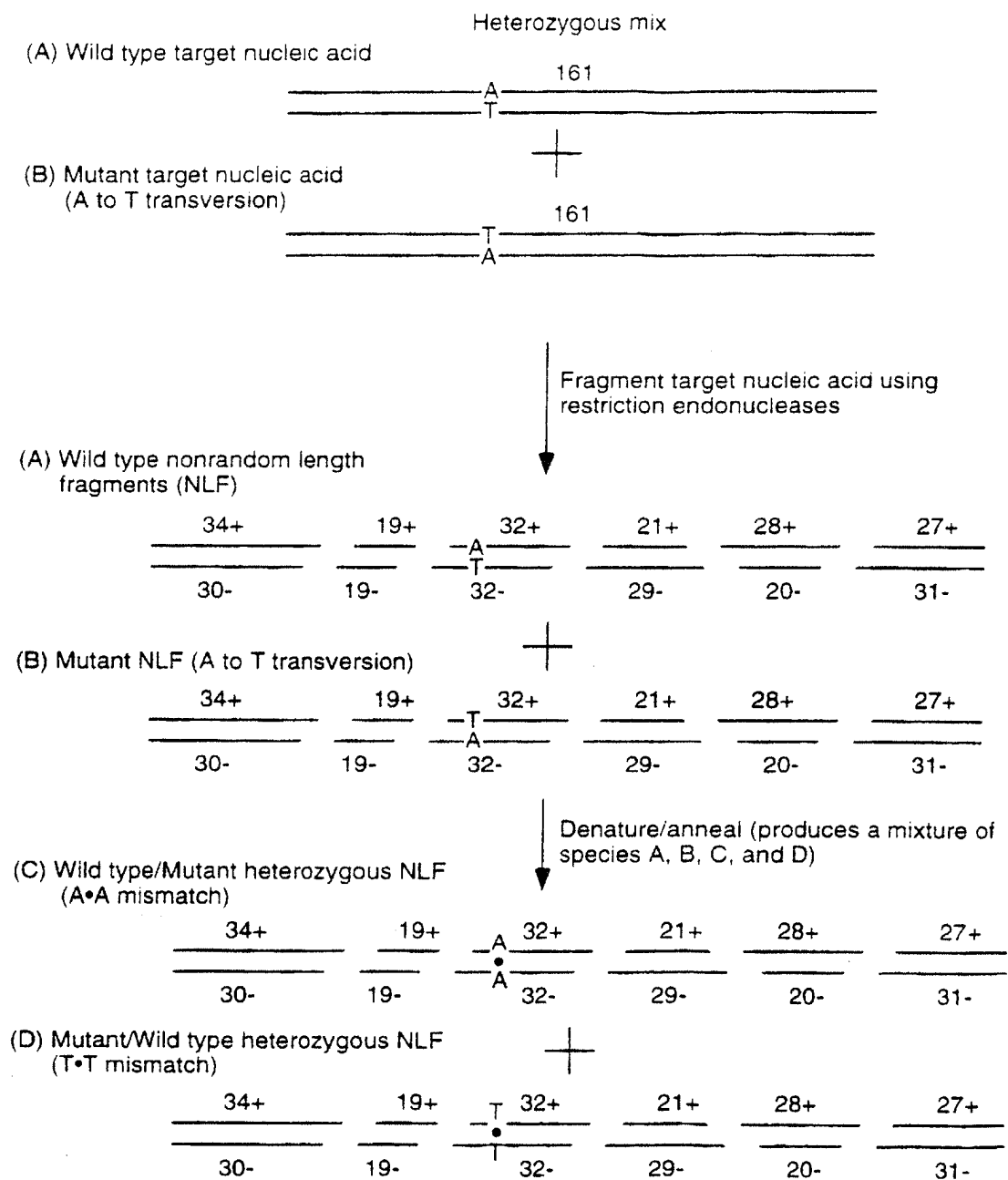
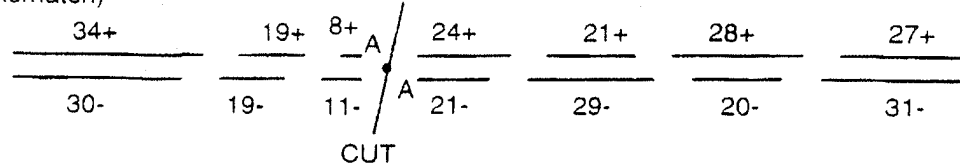


Fig. 10A

↓ mutation-specific cleavage at the location of the base mismatch (affects species (C) and (D) only).

(C) Wild type/Mutant heterozygous NLF
(A•A mismatch)



(D) Mutant/Wild type heterozygous NLF
(T•T mismatch)

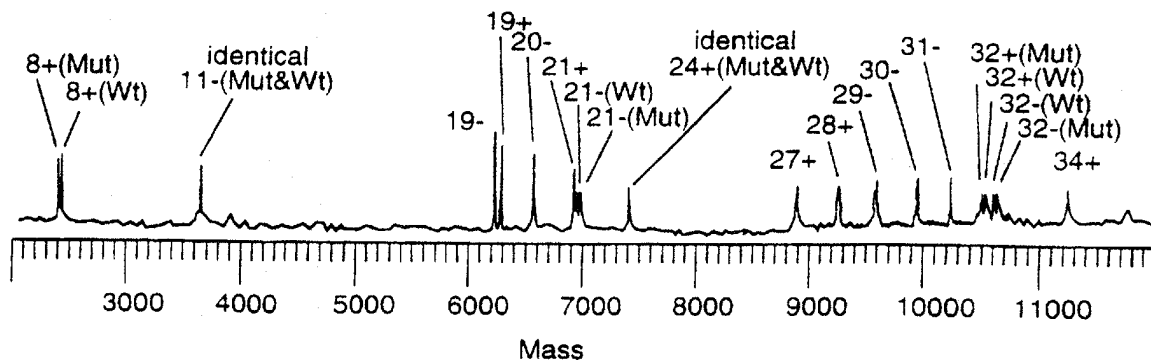
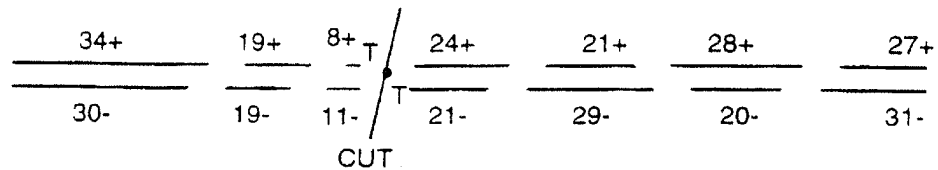


Fig. 10B

SUBSTITUTE SHEET (RULE 26)

14/22

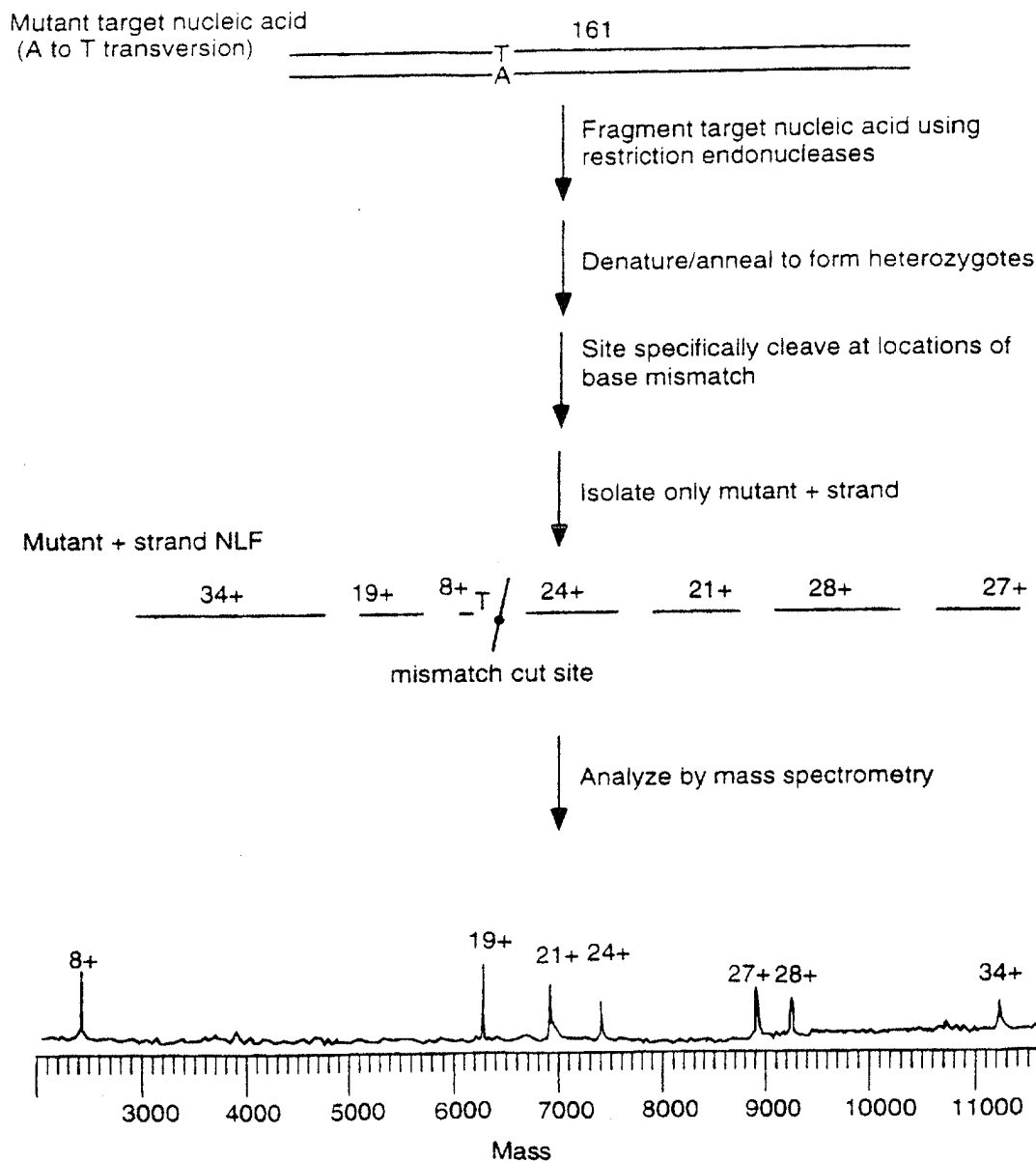


Fig. 11

SUBSTITUTE SHEET (RULE 26)

Single-stranded nucleic acid target

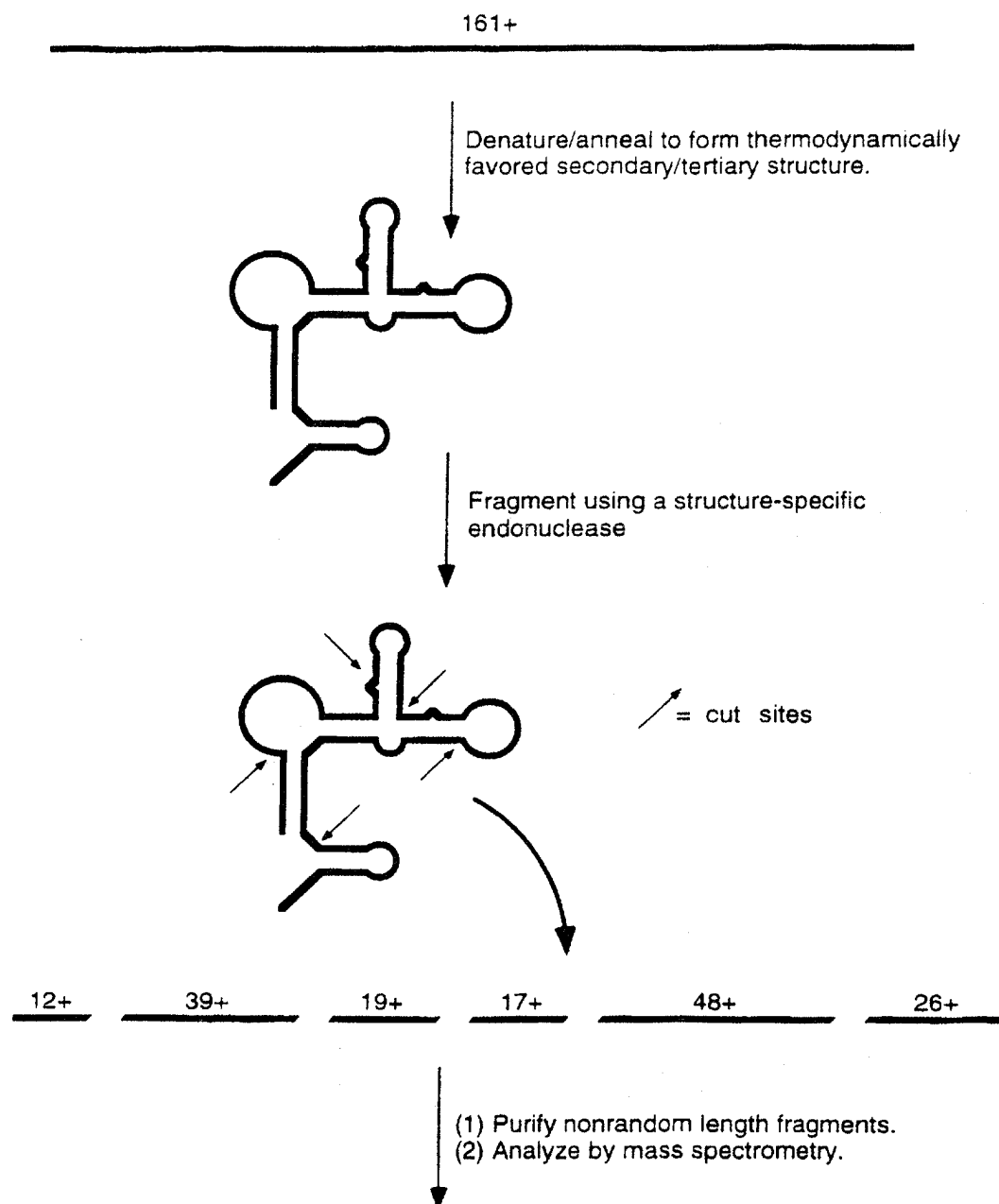
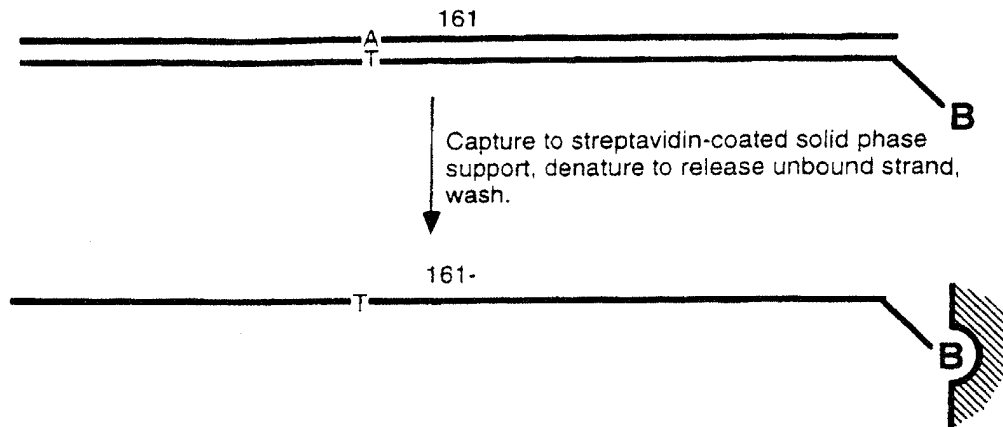


Fig. 12

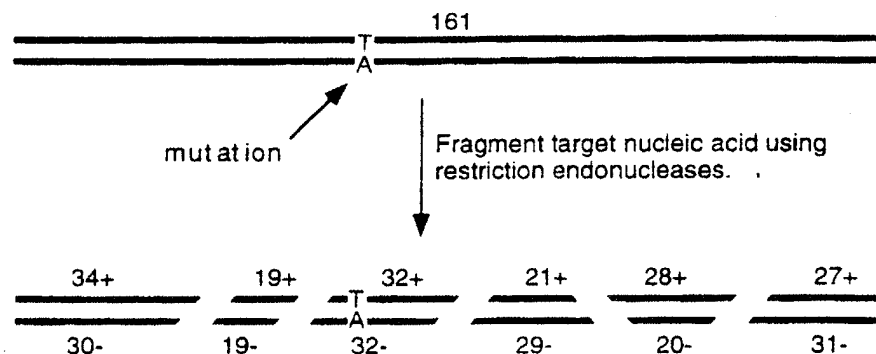
SUBSTITUTE SHEET (RULE 26)

16/22

Tube (1) Make capture probe using biotinylated primer during amplification of target.



Tube (2) Amplify target nucleic acid and fragment using restriction enzymes.



Mix contents of Tubes (1) and (2), denature/anneal + strand of fragmented target nucleic acid to solid-phase-bound capture probe.

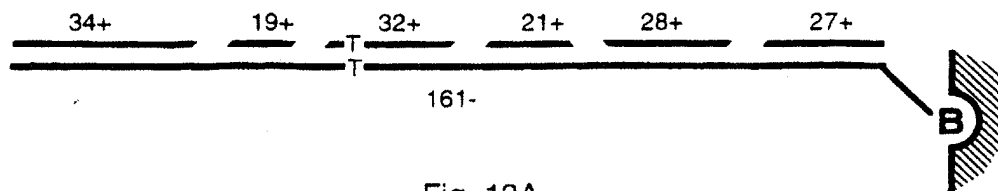


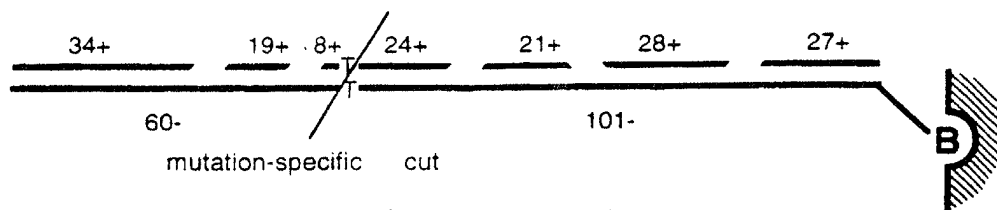
Fig. 13A

SUBSTITUTE SHEET (RULE 26)

17/22

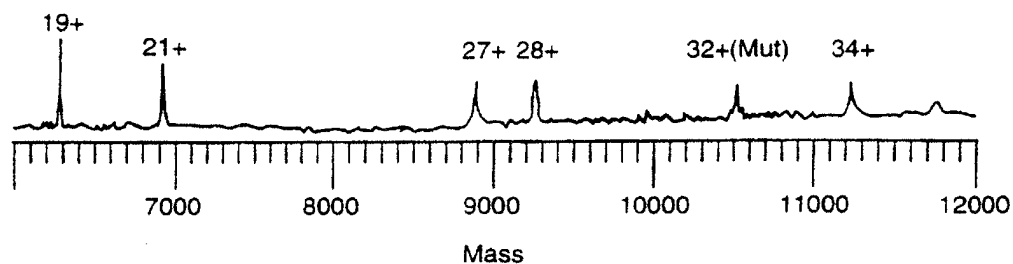
(Optional)

Cleave site-specifically at the location of any loop or mismatch using targeted endonuclease.



(1) Wash solid phase bound products to remove any unbound DNA and all contaminants. (2) Release single-stranded fragments by denaturation of the bound duplex. (3) Analyze by mass spectrometry.

Standard expected spectra



Expected spectra with optional mutation-specific cutting

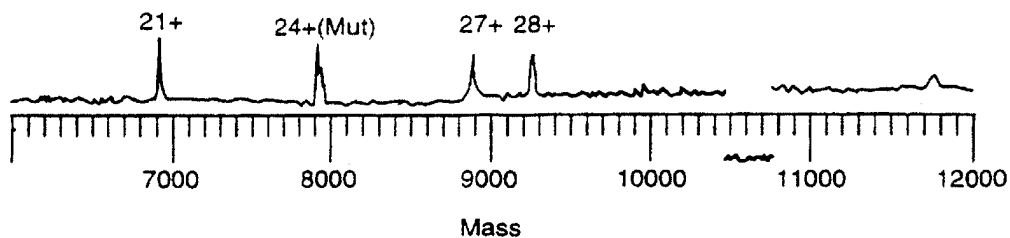


Fig. 13B

SUBSTITUTE SHEET (RULE 26)

18/22

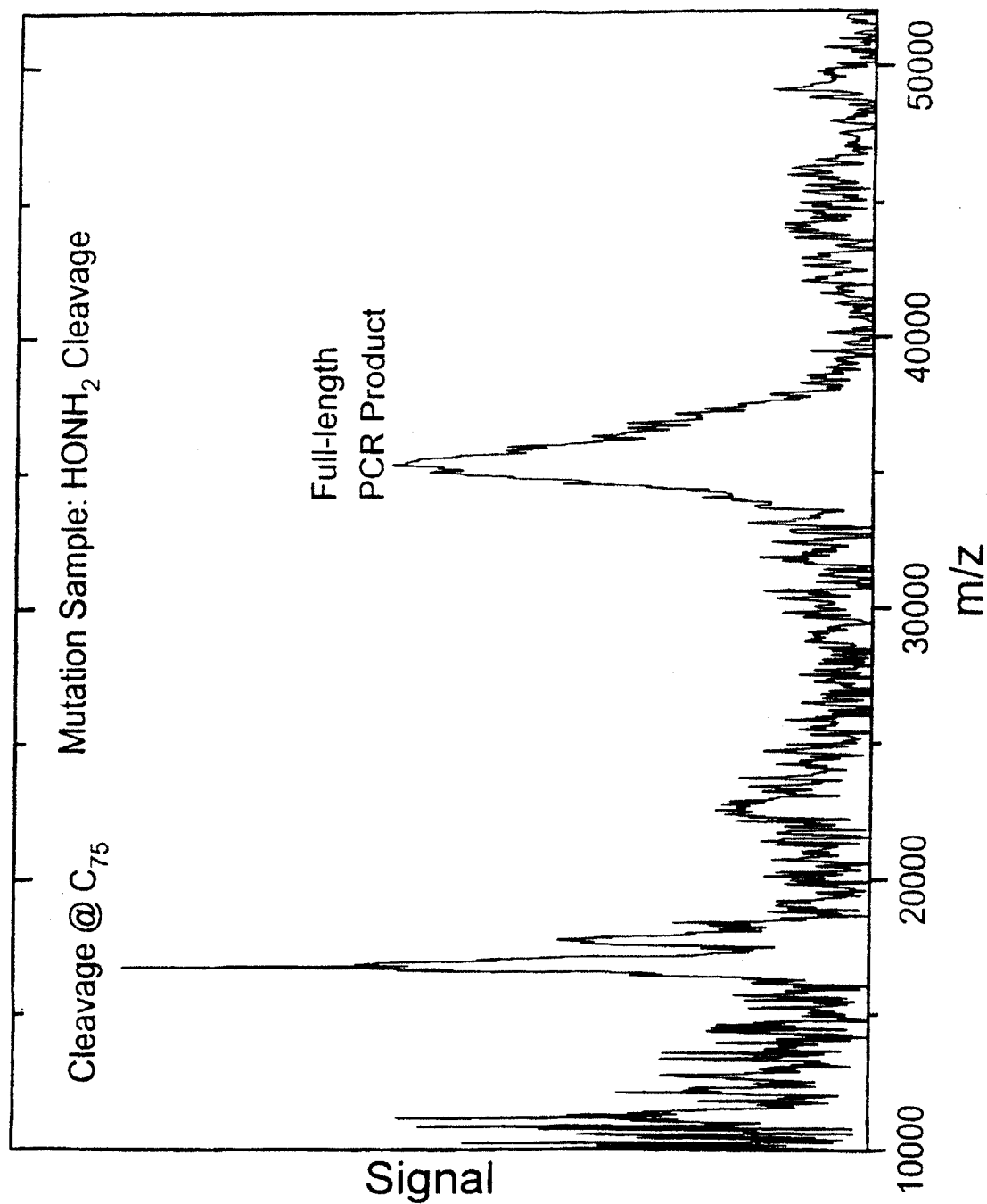


Fig. 14

SUBSTITUTE SHEET (RULE 26)

19/22

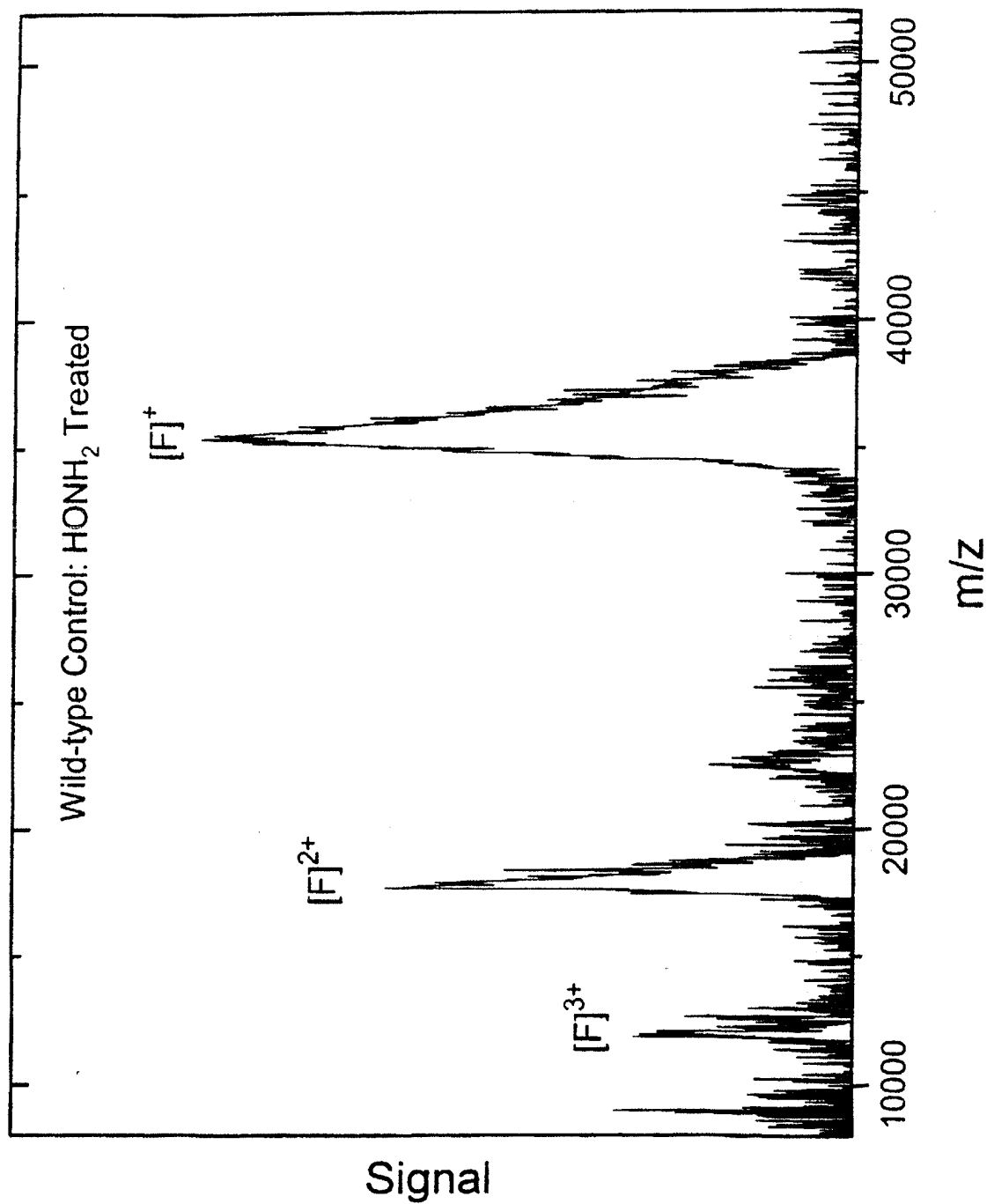


Fig. 15

SUBSTITUTE SHEET (RULE 26)

20/22

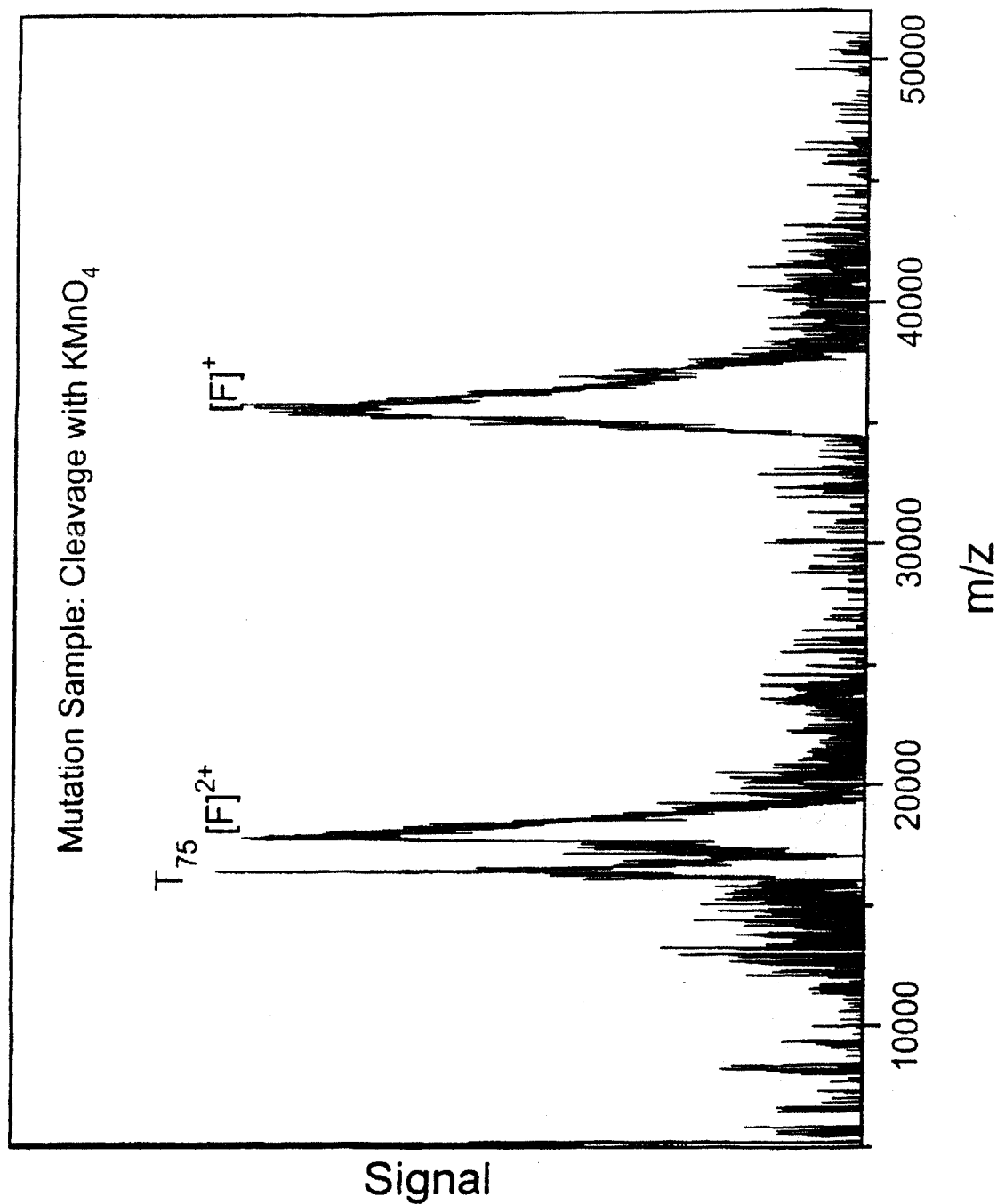


Fig. 16

SUBSTITUTE SHEET (RULE 26)

21/22

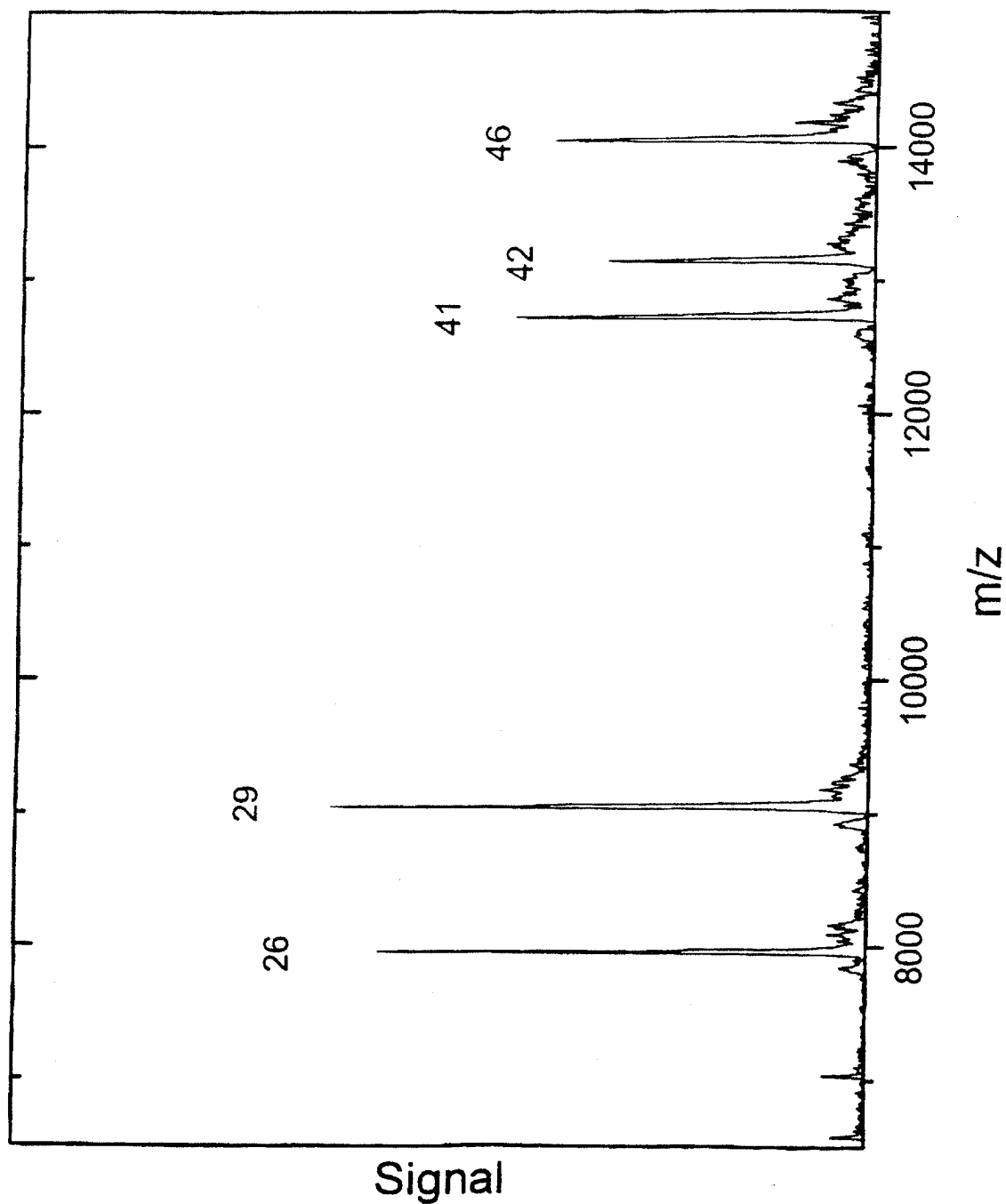
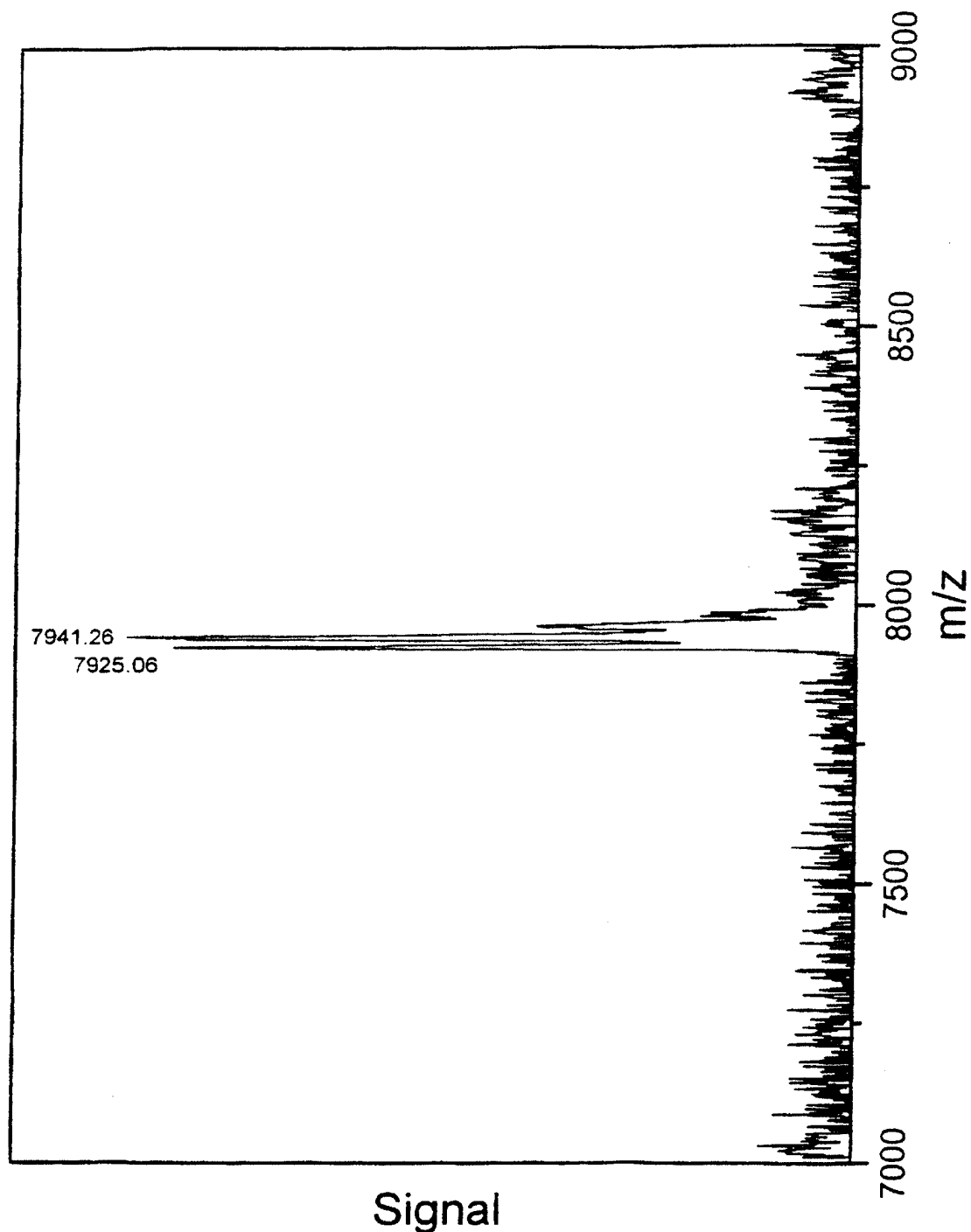


Fig. 17

SUBSTITUTE SHEET (RULE 26)

22/22



Signal

Fig. 18

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 97/03499

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ANALYTICAL CHEMISTRY, vol. 66, no. 10, 15 May 1994, pages 1637-1645, XP000579973 WU K J ET AL: "TIME-OF-FLIGHT MASS SPECTROMETRY OF UNDERIVATIZED SINGLE-STRANDED DNA OLIGOMERS BY MATRIX-ASSISTED LASER DESORPTION" see the whole document ---	1-45
Y	WO 95 07361 A (PASTEUR INSTITUT ; INST NAT SANTE RECH MED (FR); MEO TOMMASO (FR);) 16 March 1995 see the whole document ---	1-41, 43-45
Y	WO 91 15600 A (HOPE CITY) 17 October 1991 see the whole document ---	42
	--- -/-	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- * "A" document defining the general state of the art which is not considered to be of particular relevance
- * "E" earlier document but published on or after the international filing date
- * "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- * "O" document referring to an oral disclosure, use, exhibition or other means
- * "P" document published prior to the international filing date but later than the priority date claimed

- * "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- * "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- * "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- * "&" document member of the same patent family

Date of the actual completion of the international search

9 July 1997

Date of mailing of the international search report

23.07.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Müller, F

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 97/03499

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 96 32504 A (UNIV BOSTON) 17 October 1996 see abstract and claims ---	1-45
P,X	WO 96 29431 A (SEQUENOM INC) 26 September 1996 see whole document, esp. claim 48 -----	1-8, 11-40, 43-45

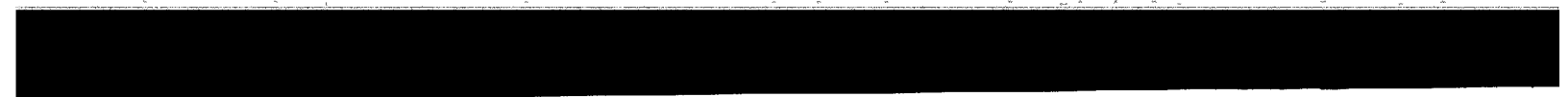
INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/03499

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9507361 A	16-03-95	FR 2709761 A CA 2171469 A EP 0717781 A	17-03-95 16-03-95 26-06-96
WO 9115600 A	17-10-91	AU 7762091 A	30-10-91
WO 9632504 A	17-10-96	AU 5544696 A	30-10-96
WO 9629431 A	26-09-96	US 5605798 A AU 5365196 A	25-02-97 08-10-96

Form PCT/ISA/210 (patent family annex) (July 1992)



1

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68		A2	(11) International Publication Number: WO 98/20166 (43) International Publication Date: 14 May 1998 (14.05.98)																						
(21) International Application Number: PCT/US97/20444 (22) International Filing Date: 6 November 1997 (06.11.97)		<p>Glendale Lake Road, Patton, PA 18668 (US). HIGGINS, G., Scott [GB/DE]; Haselweg 1, D-22880 Weidel (DE). BRAUN, Andreas [DE/US]; 13232 Benchley Road, San Diego, CA 92130 (US). DAMHOFFER-DEMAR, Brigitte [AT/US]; 3899 Haines Street #8-308, San Diego, CA 92109 (US). JURINKE, Christian [DE/DE]; Grope Hall 68, D-22115 Hamburg (DE). VAN DEN BOOM, Dirk [DE/DE]; Forsthausstrasse 8, D-63303 Preiech (DE). XIANG, Guobing [CN/US]; Apartment 23, 11381 Zapata Avenue, San Diego, CA 92126 (US). LOUGH, David, M. [GB/GB]; 32 Deanhead Road, Eyemouth, Berwickshire TD14 55A (GB).</p> <p>(74) Agent: SEIDMAN, Stephanie, L.; Brown Martin Haller & McClain, 1660 Union Street, San Diego, CA 92101-2926 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>																							
(30) Priority Data: <table border="0"><tr><td>08/744,481</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,036</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,055</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/744,590</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/786,988</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/787,639</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/933,792</td><td>19 September 1997 (19.09.97)</td><td>US</td></tr><tr><td>08/947,801</td><td>8 October 1997 (08.10.97)</td><td>US</td></tr></table> <p>(71) Applicant (for all designated States except US): SEQUENOM, INC. [US/US]; 11555 Sorrento Valley Road, San Diego, CA 92121 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): KOSTER, Hubert [DE/US]; 8636 C Via Mallorca Drive, La Jolla, CA 92037 (US). TANG, Kai [CN/US]; 8521 Summerdale Road #241, San Diego, CA 92126 (US). FU, Dong-Jing [CN/US]; 10615 Dabney Drive #21, San Diego, CA 92126 (US). SIEGERT, Carston, W. [DE/US]; Geielstrasse 42, D-22303 Hamburg (DE). LITTLE, Daniel, P. [US/US]; 393</p>				08/744,481	6 November 1996 (06.11.96)	US	08/746,036	6 November 1996 (06.11.96)	US	08/746,055	6 November 1996 (06.11.96)	US	08/744,590	6 November 1996 (06.11.96)	US	08/786,988	23 January 1997 (23.01.97)	US	08/787,639	23 January 1997 (23.01.97)	US	08/933,792	19 September 1997 (19.09.97)	US	08/947,801
08/744,481	6 November 1996 (06.11.96)	US																							
08/746,036	6 November 1996 (06.11.96)	US																							
08/746,055	6 November 1996 (06.11.96)	US																							
08/744,590	6 November 1996 (06.11.96)	US																							
08/786,988	23 January 1997 (23.01.97)	US																							
08/787,639	23 January 1997 (23.01.97)	US																							
08/933,792	19 September 1997 (19.09.97)	US																							
08/947,801	8 October 1997 (08.10.97)	US																							
(54) Title: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY																									
(57) Abstract <p>Fast and highly accurate mass spectrometry-based processes for detecting a particular nucleic acid sequence in a biological sample are provided. Depending on the sequence to be detected, the processes can be used, for example, to diagnose a genetic disease or chromosomal abnormality; a predisposition to a disease or condition, infection by a pathogenic organism, or for determining identity or heredity.</p>																									

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

-1-

DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY**Related Applications**

- For U.S. National Stage purposes, this application is a
- 5 continuation-in-part of U.S. application Serial No. 08/744,481, filed November 6, 1996, to Köster, entitled "DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY". This application is also a continuation-in-part of U.S. application Serial Nos. 08/744,590, 08/746,036, 08/746,055, 08/786,988, 08/787,639, 08/933,792 and U.S. application Serial No.
- 10 atty dkt. no. 7352-2001B, filed October 8, 1997, which is a continuation-in-part of U.S. application Nos. 08/746,055, 08/786,988 and 08/787,639. For international purposes, benefit of priority is claimed to each of these applications.

- This application is related to U.S. Patent Application Serial No.
- 15 08/617,256 filed on March 18, 1996, which is a continuation-in-part of U.S. application Serial No. 08/406,190, filed March 17, 1995, now U.S. Patent No. 5,605,798, and is also related U.S. Patent Nos. 5,547,835 and 5,622,824.

- Where permitted the subject matter of each of the above-noted
- 20 patent applications and the patent is herein incorporated in its entirety.

BACKGROUND OF THE INVENTION**Detection of mutations**

- The genetic information of all living organisms (e.g., animals, plants and microorganisms) is encoded in deoxyribonucleic acid (DNA).
- 25 In humans, the complete genome is contains of about 100,000 genes located on 24 chromosomes (The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992). Each gene codes for a specific protein, which after its expression via transcription and translation, fulfills a specific biochemical function within a living cell. Changes in a DNA
- 30 sequence are known as mutations and can result in proteins with altered

-2-

or in some cases even lost biochemical activities; this in turn can cause genetic disease. Mutations include nucleotide deletions, insertions or alterations (i.e. point mutations). Point mutations can be either "missense", resulting in a change in the amino acid sequence of a protein
5 or "nonsense" coding for a stop codon and thereby leading to a truncated protein.

More than 3000 genetic diseases are currently known (Human Genome Mutations, D. N. Cooper and M. Krawczak, BIOS Publishers, 1993), including hemophilias, thalassemias, Duchenne Muscular
10 Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF). In addition to mutated genes, which result in genetic disease, certain birth defects are the result of chromosomal abnormalities such as Trisomy 21 (Down's Syndrome), Trisomy 13 (Patau Syndrome), Trisomy 18 (Edward's Syndrome), Monosomy X
15 (Turner's Syndrome) and other sex chromosome aneuploidies such as Klienfelter's Syndrome (XXY). Further, there is growing evidence that certain DNA sequences may predispose an individual to any of a number of diseases such as diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancer (e.g., colorectal, breast, ovarian, lung).
20 Viruses, bacteria, fungi and other infectious organisms contain distinct nucleic acid sequences, which are different from the sequences contained in the host cell. Therefore, infectious organisms can also be detected and identified based on their specific DNA sequences.

Since the sequence of about 16 nucleotides is specific on
25 statistical grounds even for the size of the human genome, relatively short nucleic acid sequences can be used to detect normal and defective genes in higher organisms and to detect infectious microorganisms (e.g., bacteria, fungi, protists and yeast) and viruses. DNA sequences can even serve as a fingerprint for detection of different individuals within the

-3-

same species (see, Thompson, J. S. and M. W. Thompson, eds., Genetics in Medicine, W.B. Saunders Co., Philadelphia, PA (1991)).

- Several methods for detecting DNA are currently being used. For example, nucleic acid sequences can be identified by comparing the
- 5 mobility of an amplified nucleic acid fragment with a known standard by gel electrophoresis, or by hybridization with a probe, which is complementary to the sequence to be identified. Identification, however, can only be accomplished if the nucleic acid fragment is labeled with a sensitive reporter function (e.g., radioactive (^{32}P , ^{35}S), fluorescent or
- 10 chemiluminescent). Radioactive labels can be hazardous and the signals they produce decay over time. Non-isotopic labels (e.g., fluorescent) suffer from a lack of sensitivity and fading of the signal when high intensity lasers are being used. Additionally, performing labeling, electrophoresis and subsequent detection are laborious, time-consuming
- 15 and error-prone procedures. Electrophoresis is particularly error-prone, since the size or the molecular weight of the nucleic acid cannot be directly correlated to the mobility in the gel matrix. It is known that sequence specific effects, secondary structure and interactions with the gel matrix are causing artefacts.

20 **Use of mass spectrometry for detection and identification of nucleic acids**

- Mass spectrometry provides a means of "weighing" individual molecules by ionizing the molecules in vacuo and making them "fly" by volatilization. Under the influence of combinations of electric and magnetic fields, the ions follow trajectories depending on their individual
- 25 mass (m) and charge (z). In the range of molecules with low molecular weight, mass spectrometry has long been part of the routine physical-organic repertoire for analysis and characterization of organic molecules by the determination of the mass of the parent molecular ion. In addition, by arranging collisions of this parent molecular ion with other

-4-

particles (e.g., argon atoms), the molecular ion is fragmented forming secondary ions by the so-called collision induced dissociation (CID). The fragmentation pattern/pathway very often allows the derivation of detailed structural information. Many applications of mass spectrometric methods are known in the art, particularly in biosciences (see, e.g., Methods in Enzymol., Vol. 193: "Mass Spectrometry" (J. A. McCloskey, editor), 1990, Academic Press, New York).

Because of the apparent analytical advantages of mass spectrometry in providing high detection sensitivity, accuracy of mass measurements, detailed structural information by CID in conjunction with an MS/MS configuration and speed, as well as on-line data transfer to a computer, there has been interest in the use of mass spectrometry for the structural analysis of nucleic acids. Recent reviews summarizing this field include K.H. Schram, "Mass Spectrometry of Nucleic Acid Components, Biomedical Applications of Mass Spectrometry" 34, 203-287 (1990); and P.F. Crain, "Mass Spectrometric Techniques in Nucleic Acid Research," Mass Spectrometry Reviews 9, 505-554 (1990); see, also U.S. Patent No. 5,547,835 and U.S. Patent No. 5,622,824).

Nucleic acids, however, are very polar biopolymers that are very difficult to volatilize. Consequently, mass spectrometric detection has been limited to low molecular weight synthetic oligonucleotides for confirming an already known oligonucleotide sequence by determining the mass of the parent molecular ion, or alternatively, confirming a known sequence through the generation of secondary ions (fragment ions) via CID in an MS/MS configuration using, in particular, for the ionization and volatilization, the method of fast atomic bombardment (FAB mass spectrometry) or plasma desorption (PD mass spectrometry). As an example, the application of FAB to the analysis of protected dimeric blocks for chemical synthesis of oligodeoxynucleotides has been

-5-

described (Köster *et al.* (1987) *Biomed. Environ. Mass Spectrometry* **14**, 111-116).

Other ionization/desorption techniques include electrospray/ion-spray (ES) and matrix-assisted laser desorption/ionization (MALDI). ES
5 mass spectrometry has been introduced by Fenn *et al.* (*J. Phys. Chem.* **88**:4451-59 (1984); PCT Application No. WO 90/14148) and current applications are summarized in review articles (see, *e.g.*, Smith *et al.* (1990) *Anal. Chem.* **62**:882-89 and Ardrey (1992) *Electrospray Mass Spectrometry, Spectroscopy Europe* **4**:10-18). The molecular weights of
10 a tetradecanucleotide (see, Covey *et al.* (1988) The "Determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ionspray Mass Spectrometry," *Rapid Commun. in Mass Spectrometry* **2**:249-256), and of a 21-mer (*Methods in Enzymol.*, **193**, "Mass Spectrometry" (McCloskey, editor), p. 425, 1990, Academic Press, New York) have
15 been published. As a mass analyzer, a quadrupole is most frequently used. Because of the presence of multiple ion peaks that all could be used for the mass calculation, the determination of molecular weights in femtomole amounts of sample is very accurate.

MALDI mass spectrometry, in contrast, can be attractive when a
20 time-of-flight (TOF) configuration (see, Hillenkamp *et al.* (1990) pp 49-60 in "Matrix Assisted UV-Laser Desorption/Ionization: A New Approach to Mass Spectrometry of Large Biomolecules," *Biological Mass Spectrometry*, Burlingame and McCloskey, editors, Elsevier Science Publishers, Amsterdam) is used as a mass analyzer. Since, in most
25 cases, no multiple molecular ion peaks are produced with this technique, the mass spectra, in principle, look simpler compared to ES mass spectrometry.

-6-

Although DNA molecules up to a molecular weight of 410,000 daltons have been desorbed and volatilized (Williams *et al.*, "Volatilization of High Molecular Weight DNA by Pulsed Laser Ablation of Frozen Aqueous Solutions," Science 246, 1585-87 (1989)), this technique had

5 only shown very low resolution (oligothymidylic acids up to 18 nucleotides, Huth-Fehre *et al.* Rapid Commun. in Mass Spectrom., 6, 209-13 (1992); DNA fragments up to 500 nucleotides in length K. Tang *et al.*, Rapid Commun. in Mass Spectrom., 8, 727-730 (1994); and a double-stranded DNA of 28 base pairs (Williams *et al.*, "Time-of-Flight

10 Mass Spectrometry of Nucleic Acids by Laser Ablation and Ionization from a Frozen Aqueous Matrix," Rapid Commun. in Mass Spectrom., 4, 348-351 (1990)). Japanese Patent No. 59-131909 describes an instrument, which detects nucleic acid fragments separated either by electrophoresis, liquid chromatography or high speed gel filtration. Mass

15 spectrometric detection is achieved by incorporating into the nucleic acids, atoms, such as S, Br, I or Ag, Au, Pt, Os, Hg, that normally do not occur in DNA.

Co-owned U.S. Patent No. 5,622,824 describes methods for DNA sequencing based on mass spectrometric detection. To achieve this, the

20 DNA is by means of protection, specificity of enzymatic activity, or immobilization, unilaterally degraded in a stepwise manner via exonuclease digestion and the nucleotides or derivatives detected by mass spectrometry. Prior to the enzymatic degradation, sets of ordered deletions that span a cloned DNA fragment can be created. In this

25 manner, mass-modified nucleotides can be incorporated using a combination of exonuclease and DNA/RNA polymerase. This permits either multiplex mass spectrometric detection, or modulation of the activity of the exonuclease so as to synchronize the degradative process.

-7-

Co-owned U.S. Patent Nos. 5,605,798 and 5,547,835 provide methods for detecting a particular nucleic acid sequence in a biological sample. Depending on the sequence to be detected, the processes can be used, for example, in methods of diagnosis. These methods, while
5 broadly useful and applicable to numerous embodiments, represent the first disclosure of such applications and can be improved upon.

Therefore, it is an object herein to provide improved methods for sequencing and detecting DNA molecules in biological samples. It is also an object herein to provide improved methods for diagnosis of genetic
10 diseases, predispositions to certain diseases, cancers, and infections.

SUMMARY OF THE INVENTION

Methods of diagnosis by detecting and/or determining sequences of nucleic acids that are based on mass spectrometry are provided herein. Methods are provided for detecting double-stranded DNA, detecting
15 mutations and other diagnostic markers using MS analysis. In particular, methods for diagnosing neuroblastoma, detecting heredity relationships, HLA compatibility, genetic fingerprinting, detecting telomerase activity for cancer diagnosis are provided.

In certain embodiments the DNA is immobilized on a solid support
20 either directly or via a linker and/or bead. Three permutations of the methods for DNA detection in which immobilized DNA is used are exemplified. These include: (1) immobilization of a template; hybridization of the primer; extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the
25 primer and Endonuclease degradation (sequencing); (2) immobilization of a primer; hybridization of a single stranded template; and extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the primer and Endonuclease degradation (sequencing); (3) immobilization of the primer; hybridization of a double

-8-

stranded template; extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the primer and Endonuclease degradation (sequencing).

In certain embodiments the DNA is immobilized on the support via
5 a selectively cleavable linker. Selectively cleavable linkers include, but are not limited to photocleavable linkers, chemically cleavable linkers and an enzymatically (such as a restriction site (nucleic acid linker), a protease site) cleavable linkers. Inclusion of a selectively cleavable linker expands the capabilities of the MALDI-TOF MS analysis because it allows
10 for all of the permutations of immobilization of DNA for MALDI-TOF MS, the DNA linkage to the support through the 3'- or 5'-end of a nucleic acid; allows the amplified DNA or the target primer to be extended by DNA synthesis; and further allows for the mass of the extended product (or degraded product via exonuclease degradation) to be of a size that is
15 appropriate for MALDI-TOF MS analysis (i.e., the isolated or synthesized DNA can be large and a small primer or a large primer sequence can be used and a small restriction fragment of a gene or single strand thereof hybridized thereto).

In a preferred embodiment, the selectively cleavable linker is a
20 chemical or photocleavable linker that is cleaved during the ionizing step of mass spectrometry. Exemplary linkers include linkers containing, a disulfide group, a leuvinyl group, an acid-labile trityl group and a hydrophobic trityl group. In other embodiments, the enzymatically cleavable linker can be a nucleic acid that is an RNA nucleotide or that
25 encodes a restriction endonuclease site. Other enzymatically cleavable linkers include linkers that contain a pyrophosphate group, an arginine-arginine group and a lysine-lysine group. Other linkers are exemplified herein.

-9-

- Methods for sequencing long fragments of DNA are provided. To perform such sequencing, specific base terminated fragments are generated from a target nucleic acid. The analysis of fragments rather than the full length nucleic acid shifts the mass of the ions to be
- 5 determined into a lower mass range, which is generally more amenable to mass spectrometric detection. For example, the shift to smaller masses increases mass resolution, mass accuracy and, in particular, the sensitivity for detection. Hybridization events and the actual molecular weights of the fragments as determined by mass spectrometry provide
- 10 sequence information (e.g., the presence and/or identity of a mutation). In a preferred embodiment, the fragments are captured on a solid support prior to hybridization and/or mass spectrometry detection. In another preferred embodiment, the fragments generated are ordered to provide the sequence of the larger nucleic acid.
- 15 One preferred method for generating base specifically terminated fragments from a nucleic acid is effected by contacting an appropriate amount of a target nucleic acid with an appropriate amount of a specific endonuclease, thereby resulting in partial or complete digestion of the target nucleic acid. Endonucleases will typically degrade a sequence into
- 20 pieces of no more than about 50-70 nucleotides, even if the reaction is not run to full completion. In a preferred embodiment, the nucleic acid is a ribonucleic acid and the endonuclease is a ribonuclease (RNase) selected from among: the G-specific RNase T₁, the A-specific RNase U₂, the A/U specific RNase PhyM, U/C specific RNase A, C specific chicken
- 25 liver RNase (RNase CL3) or crisavitin. In another preferred embodiment, the endonuclease is a restriction enzyme that cleaves at least one site contained within the target nucleic acid. Another preferred method for generating base specifically terminated fragments includes performing a combined amplification and base-specific termination reaction (e.g., using

-10-

- an appropriate amount of a first DNA polymerase, which has a relatively low affinity towards the chain-terminating nucleotides resulting in an exponential amplification of the target; and a polymerase with a relatively high affinity for the chain terminating nucleotide resulting in base-specific
- 5 termination of the polymerization. Inclusion of a tag at the 5' and/or 3' end of a target nucleic acid can facilitates the ordering of fragments.

- Methods for determining the sequence of an unknown nucleic acid in which the 5' and/or 3' end of the target nucleic acid can include a tag are provided. Inclusion of a non-natural tag on the 3' end is also useful
- 10 for ruling out or compensating for the influence of 3' heterogeneity, premature termination and nonspecific elongation. In a preferred embodiment, the tag is an affinity tag (e.g., biotin or a nucleic acid that hybridizes to a capture nucleic acid). Most preferably the affinity tag facilitates binding of the nucleic acid to a solid support. In another
- 15 preferred embodiment, the tag is a mass marker (i.e. a marker of a mass that does not correspond to the mass of any of the four nucleotides). In a further embodiment, the tag is a natural tag, such as a polyA tail or the natural 3' heterogeneity that can result, for example, from a transcription reaction.

- 20 Methods of sequence analysis in which nucleic acids have been replicated from a nucleic acid molecule obtained from a biological sample are specifically digested using one or more nucleases (deoxyribonucleases for DNA, and ribonucleases for RNA) are provided. The fragments captured on a solid support carrying the corresponding
- 25 complementary sequences. Hybridization events and the actual molecular weights of the captured target sequences provide information on mutations in the gene. The array can be analyzed spot-by-spot using mass spectrometry. Further, the fragments generated can be ordered to provide the sequence of the larger target fragment.

-11-

In another embodiment, at least one primer with a 3'-terminal base is hybridized to the target nucleic acid near a site where possible mutations are to be detected. An appropriate polymerase and a set of three nucleoside triphosphates (NTPs) and the fourth added as a
5 terminator are reacted. The extension reaction products are measured by mass spectrometry and are indicative of the presence and the nature of a mutation. The set of three NTPs and one dd-NTP (or three NTPs and one 3'-deoxy NTP), will be varied to be able to discriminate between several mutations (including compound heterozygotes) in the target nucleic acid
10 sequence.

Methods for detecting and diagnosing neoplasia/malignancies in a tissue or cell sample are provided. The methods rely on a telomeric repeat amplification protocol (TRAP) -MS assay and include the steps of:

- 15 a) obtaining a tissue or a cell sample, such as a clinical isolate or culture of suspected cells;
- b) isolating/extracting/purifying telomerase from the sample;
- c) adding the telomerase extract to a composition
20 containing a synthetic DNA primer, which is optionally immobilized, complementary to the telomeric repeat, and all four dNTPs under conditions that result in telomerase specific extension of the synthetic DNA;
- 25 d) amplifying the telomerase extended DNA products,, preferably using a primer that contains a "linker moiety", such as a moiety based on thiol chemistry or streptavidin;

-12-

- 5
- e) isolating linker-amplified primers, such as by using a complementary binding partner immobilized on a solid support;
 - f) optionally conditioning the DNA for crystal formation; and
 - g) performing MS by ionizing/volatizing the sample to detect the DNA product.

Telomerase-specific extension is indicative of neoplasia/malignancy.

This method can be used to detect specific malignancies. The use
10 of MS to detect the DNA product permits identification the extended product, which is indicative of telomerase activity in the sample.
If desired, the synthetic DNA can be in the form an array.

Methods for detecting mutations are provided and the use thereof
oncogenes and to thereby screen for transformed cells, which are
15 indicative of neoplasia. Detection of mutations present in oncogenes are indicative of transformation. This method includes the steps of:

- a) obtaining a biological sample;
- b) amplifying a portion of the selected proto-oncogene that includes a codon indicative of transformation,
20 where one primer has a linker moiety for immobilization;
- c) immobilizing DNA via the linker moiety to a solid support, optionally in the form of an array;
- d) hybridizing a primer complementary to the proto
25 oncogene sequence that is upstream from the codon
 - e) adding 3dNTPs/1 ddNTP and DNA polymerase and extending the hybridized primer to the next ddNTP location;

-13-

- f) ionizing/volatizing the sample; and
 - g) detecting the mass of the extended DNA, whereby mass indicates the presence of wild-type or mutant alleles. The presence of a mutant allele at the codon
- 5 is diagnostic for neoplasia.

In an exemplary embodiment, extension-MS analysis is used detect the presence of a mutated codon 634 in the retrovirus (RET)-proto oncogene.

- In another embodiment, methods for diagnosing diseases using reverse transcription and amplification of a gene expressed in
- 10 transformed cells. In particular, a method for diagnosis of neuroblastoma using reverse transcriptase (RT)-MS of tyrosine hydroxylase, which is a catecholamine biosynthetic enzyme that expressed in tumor cells, but not in tumor cells but not normal cells, such as normal bone marrow cells is provided. The method includes the steps
- 15 of:

- a) obtaining a tissue sample;
 - b) isolating polyA RNA from the sample;
 - c) preparing a cDNA library using reverse transcription;
 - d) amplifying a cDNA product, or portion thereof, of the
- 20 selected gene, where one oligo primer has a linker moiety;
- e) isolating the amplified product by immobilizing the DNA to solid support via the linker moiety;
 - f) optionally conditioning the DNA:
- 25 g) ionizing/volatizing sample and detecting the presence of a DNA peak that is indicative of expression of the selected gene gene. For example, expression of the tyrosine hydroxylase gene is indicative of neuroblastoma.

-14-

Also provided are methods of directly detecting a double-stranded nucleic acid using MALDI-TOF MS. These methods include the steps of:

- 5 a) isolating a double stranded DNA of an appropriate size for MS via amplification methods or formed by hybridization of single-stranded DNA fragment;
- b) preparing the double-stranded DNA for analysis under conditions that increase the ratio of dsDNA:ssDNA in which the conditions include one or all of the following: preparing samples for analysis at reduced temperatures (i.e. 4 ° C),
10 and using of higher DNA concentrations in the matrix to drive duplex formation
- c) ionizing/volatizing the sample of step b), where this step uses low acceleration voltage of the ions to assist in maintaining duplex DNA by, for example, adjusting laser power to just above threshold irradiation for ionization, and
15 d) detecting the presence of the dsDNA of the appropriate mass.

In preferred embodiments, the matrix includes 3-hydroxypicolinic acid. The detected DNA can be indicative of a genetic disorder, genetic
20 disease, genetic predisposition to a disease chromosomal abnormalities. In other embodiments, the mass of the double stranded DNA is indicative of the deletion, insertion, mutation.

A method designated primer oligo base extension (PROBE) is provided. This method uses a single detection primer followed by an
25 oligonucleotide extension step to give products, which can be readily resolved by MALDI-TOF mass spectrometry. The products differ in length by a number of bases specific for a number of repeat units or for second site mutations within the repeated region. The method is exemplified using as a model system the AluVpA polymorphism in intron

-15-

5 of the interferon- α receptor gene located on human chromosome 21, and the poly T tract of the splice acceptor site of intron 8 from the CFTR gene located on human chromosome 7. The method is advantageously used for example, for determining identity, identifying mutations, familial relationship, HLA compatability and other such markers ,using PROBE-MS analysis of microsatellite DNA. In a preferred embodiment, the method includes the steps of:

- a) obtaining a biological sample from two individuals;
- b) amplifying a region of DNA from each individual that
10 contains two or more microsatellite DNA repeat sequences
- c) ionizing/volatizing the amplified DNA;
- d) detecting the presence of the amplified DNA and comparing
15 the molecular weight of the amplified DNA. Different sizes are indicative of non-identity (i.e. wild-type versus mutation), non-heredity or non-compatibility; similar size fragments indicate the possibility identity, of familial relationship, or HLA compatibility.

More than one marker may be examined simulataneoulsy, primers with different linker moieties are used for immobilization.

20 Another method loop-primer oligo base extension, designated LOOP-PROBE, for detection of mutations especially predominant disease causing mutations or common polymorphisms is provided. In a particular embodiment, this method for detecting target nucleic acid in a sample, includes the steps of:

- 25 a) amplifying a target nucleic acid sequence, such as β -globin, in a sample, using (i) a first primer whose 5'-end shares identity to a portion of the target DNA immediately downstream from the targeted codon followed by a sequence that introduces a unique restriction endonuclease

-16-

- 5 site, such as CfoI in the case of β -globin, into the amplicon and whose 3'-end primer is self-complementary; and (ii) a second downstream primer that contains a tag, such as biotin, for immobilizing the DNA to a solid support, such as streptavidin beads;
- c) immobilizing the double-stranded amplified DNA to a solid support via the linker moiety;
- d) denaturing the immobilized DNA and isolating the non-immobilized DNA strand;
- 10 e) annealing the intracomplementary sequences in the 3'-end of the isolated non-immobilized DNA strand, such that the 3'-end is extendable by a polymerase, which annealing can be performed, for example, by heating then and cooling to about 37° C, or other suitable method;
- 15 f) extending the annealed DNA by adding DNA polymerase, 3 dNTPs/1 ddNTP, whereby the 3'-end of the DNA strand is extended by the DNA polymerase to the position of the next ddNTP location (i.e., to the mutation location);
- g) cleaving the extended double stranded stem loop DNA with the unique restriction endonuclease and removing the
- 20 cleaved stem loop DNA
- i) (optionally adding a matrix) ionizing/volatizing the extended product; and
- j) detecting the presence of the extended target nucleic acid,
- 25 whereby the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation at the target codon(s).

This method eliminates one specific reagent for mutation detection compared other methods of MS mutational analyses, thereby simplifying

-17-

the process and rendering it amenable to automation. Also, the specific extended product that is analyzed is cleaved from the primer and is therefore shorter compared to the other methods. In addition, the annealing efficiency is higher compared to annealing of an added primer and should therefore generate more product. The process is compatible with multiplexing and various detection schemes (e.g., single base extension, oligo base extension and sequencing). For example, the extension of the loop-primer can be used for generation of short diagnostic sequencing ladders within highly polymorphic regions to perform, for example, HLA typing or resistance as well as species typing.

In another embodiment, a method of detecting a target nucleic acid in a biological sample using RNA amplification is provided. In the method, the target is amplified the target nucleic acid, using a primer that shares a region complementary to the target sequence and upstream encodes a promoter, such as the T7 promoter. A DNA-dependent RNA polymerase and appropriate ribonucleotides are added to synthesize RNA, which is analyzed by MS.

Improved methods of sequencing DNA using MS are provided. In these methods thermocycling for amplification is used prior to MS analysis, thereby increasing the signal.

Also provided are primers for use in MS analyses. In particular, primers, comprising all or, for longer oligonucleotides, at least about 20, preferably about 16, bases of any of the sequence of nucleotides sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95, 98, 101-110, 112-123, 126, 128, 129, and primers set forth in SEQ ID Nos. 280-287. The primers are unlabeled, and optionally include a mass modifying moiety, which is preferably attached to the 5' end.

-18-

Other features and advantages of the methods provided herein will be further described with reference to the following Figures, Detailed Description and Claims.

BRIEF DESCRIPTION OF THE FIGURES

- 5 FIGURE 1A is a diagram showing a process for performing mass spectrometric analysis on one target detection site (TDS) contained within a target nucleic acid molecule (T), which has been obtained from a biological sample. A specific capture sequence (C) is attached to a solid support (SS) via a spacer (S). The capture sequence is chosen to
- 10 specifically hybridize with a complementary sequence on the target nucleic acid molecule (T), known as the target capture site (TCS). The spacer (S) facilitates unhindered hybridization. A detector nucleic acid sequence (D), which is complementary to the TDS is then contacted with the TDS. Hybridization between D and the TDS can be detected by mass
- 15 spectrometry.

- FIGURE 1B is a diagram showing a process for performing mass spectrometric analysis on at least one target detection site (here TDS 1 and TDS 2) via direct linkage to a solid support. The target sequence (T) containing the target detection site (TDS 1 and TDS 2) is immobilized to
- 20 a solid support via the formation of a reversible or irreversible bond formed between an appropriate functionality (L') on the target nucleic acid molecule (T) and an appropriate functionality (L) on the solid support. Detector nucleic acid sequences (here D1 and D2), which are complementary to a target detection site (TDS 1 or TDS 2) are then
- 25 contacted with the TDS. Hybridization between TDS 1 and D1 and/or TDS 2 and D2 can be detected and distinguished based on molecular weight differences.

FIGURE 1C is a diagram showing a process for detecting a wildtype (D^{wt}) and/or a mutant (D^{mut}) sequence in a target (T) nucleic acid

-19-

molecule. As in Figure 1A, a specific capture sequence (C) is attached to a solid support (SS) via a spacer (S). In addition, the capture sequence is chosen to specifically interact with a complementary sequence on the target sequence (T), the target capture site (TCS) to be detected through hybridization. If the target detection site (TDS) includes a mutation, X, detection sites can be distinguished from wildtype by mass spectrometry. Preferably, the detector nucleic acid molecule (D) is designed so that the mutation is in the middle of the molecule and therefore would not lead to a stable hybrid if the wildtype detector oligonucleotide (D^{wt}) is contacted with the target detector sequence, e.g., as a control. The mutation can also be detected if the mutated detector oligonucleotide (D^{mut}) with the matching base at the mutated position is used for hybridization. If a nucleic acid molecule obtained from a biological sample is heterozygous for the particular sequence (i.e. contain D^{wt} and D^{mut}), D^{wt} and D^{mut} will be bound to the app and D^{mut} to be detected simultaneously.

FIGURE 2 is a diagram showing a process in which several mutations are simultaneously detected on one target sequence molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide.

FIGURE 3 is a diagram showing still another multiplex detection format. In this embodiment, differentiation is accomplished by employing different specific capture sequences which are position-specifically immobilized on a flat surface (e.g., a 'chip array'). If different target sequences T1-Tn are present, their target capture sites TCS1-TCSn will interact with complementary immobilized capture sequences C1-Cn. Detection is achieved by employing appropriately mass differentiated

-20-

detector oligonucleotides D1-Dn, which are mass differentiated either by their sequences or by mass modifying functionalities M1-Mn.

FIGURE 4 is a diagram showing a format wherein a predesigned target capture site (TCS) is incorporated into the target sequence using
5 nucleic acid (i.e., PCR) amplification. Only one strand is captured, the other is removed (e.g., based on the interaction between biotin and streptavidin coated magnetic beads). If the biotin is attached to primer 1 the other strand can be appropriately marked by a TCS. Detection is as described above through the interaction of a specific detector
10 oligonucleotide D with the corresponding target detection site TDS via mass spectrometry.

FIGURE 5 is a diagram showing how amplification (here ligase chain reaction (LCR)) products can be prepared and detected by mass spectrometry. Mass differentiation can be achieved by the mass
15 modifying functionalities (M1 and M2) attached to primers (P1 and P4 respectively). Detection by mass spectrometry can be accomplished directly (i.e. without employing immobilization and target capturing sites (TCS)). Multiple LCR reactions can be performed in parallel by providing an ordered array of capturing sequences (C). This format allows
20 separation of the ligation products and spot by spot identification via mass spectrometry or multiplexing if mass differentiation is sufficient.

FIGURE 6A is a diagram showing mass spectrometric analysis of a nucleic acid molecule, which has been amplified by a transcription amplification procedure. An RNA sequence is captured via its TCS
25 sequence, so that wildtype and mutated target detection sites can be detected as above by employing appropriate detector oligonucleotides (D).

-21-

FIGURE 6B is a diagram showing multiplexing to detect two different (mutated) sites on the same RNA in a simultaneous fashion using mass-modified detector oligonucleotides M1-D1 and M2-D2.

FIGURE 6C is a diagram of a different multiplexing procedure for
5 detection of specific mutations by employing mass modified dideoxynucleoside or 3'-deoxynucleoside triphosphates and an RNA dependent DNA polymerase. Alternatively, DNA dependent RNA polymerase and ribonucleotide phosphates can be employed. This format allows for simultaneous detection of all four base possibilities at the site
10 of a mutation (X).

FIGURE 7A is a diagram showing a process for performing mass spectrometric analysis on one target detection site (TDS) contained within a target nucleic acid molecule (T), which has been obtained from a biological sample. A specific capture sequence (C) is attached to a solid
15 support (SS) via a spacer (S). The capture sequence is chosen to specifically hybridize with a complementary sequence on T known as the target capture site (TCS). A nucleic acid molecule that is complementary to a portion of the TDS is hybridized to the TDS 5' of the site of a mutation (X) within the TDS. The addition of a complete set of
20 dideoxynucleosides or 3'-deoxynucleoside triphosphates (e.g., pppAdd, pppTdd, pppCdd and pppGdd) and a DNA dependent DNA or RNA polymerase allows for the addition only of the one dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to X.

FIGURE 7B is a diagram showing a process for performing mass
25 spectrometric analysis to determine the presence of a mutation at a potential mutation site (M) within a nucleic acid molecule. This format allows for simultaneous analysis of alleles (A) and (B) of a double stranded target nucleic acid molecule, so that a diagnosis of homozygous normal, homozygous mutant or heterozygous can be provided. Allele A

-22-

and B are each hybridized with complementary oligonucleotides ((C) and (D) respectively), that hybridize to A and B within a region that includes M. Each heteroduplex is then contacted with a single strand specific endonuclease, so that a mismatch at M, indicating the presence of a
5 mutation, results in the cleavage of (C) and/or (D), which can then be detected by mass spectrometry.

FIGURE 8 is a diagram showing how both strands of a target DNA can be prepared for detection using transcription vectors having two different promoters at opposite locations (e.g., the SP6 and T7
10 promoter). This format is particularly useful for detecting heterozygous target detections sites (TDS). Employing the SP6 or the T7 RNA polymerase both strands could be transcribed separately or simultaneously. The transcribed RNA molecules can be specifically captured and simultaneously detected using appropriately mass-
15 differentiated detector oligonucleotides. This can be accomplished either directly in solution or by parallel processing of many target sequences on an ordered array of specifically immobilized capturing sequences.

FIGURE 9 is a diagram showing how RNA prepared as described in Figures 6, 7 and 8 can be specifically digested using one or more
20 ribonucleases and the fragments captured on a solid support carrying the corresponding complementary sequences. Hybridization events and the actual molecular weights of the captured target sequences provide information on whether and where mutations in the gene are present. The array can be analyzed spot by spot using mass spectrometry. DNA
25 can be similarly digested using a cocktail of nucleases including restriction endonucleases. Mutations can be detected by different molecular weights of specific, individual fragments compared to the molecular weights of the wildtype fragments.

-23-

FIGURE 10A shows UV spectra resulting from the experiment described in the following Example 1. Panel i) shows the absorbance of the 26-mer before hybridization. Panel ii) shows the filtrate of the centrifugation after hybridization. Panel iii) shows the results after the first wash with 50 mM ammonium citrate. Panel iv) shows the results after the second wash with 50 mM ammonium citrate.

FIGURE 10B shows a mass spectrum resulting from the experiment described in the following Example 1 after three washing/centrifugation steps.

FIGURE 10C shows a mass spectrum resulting from the experiment described in the following Example 1 showing the successful desorption of the hybridized 26-mer off of beads in accordance with the format depicted schematically in Figure 1B.

FIGURE 11 shows a mass spectrum resulting from the experiment described in the following Example 1 showing the giving proof of an experiment as schematically depicted in FIGURE 1B successful desorption of the hybridized 40-mer. The efficiency of detection suggests that fragments much longer than 40-mers can also be desorbed. Figure 12 shows a mass spectrum resulting from the experiment described in the following Example 2 showing the successful desorption and differentiation of an 18-mer and 19-mer by electrospray mass spectrometry, the mixture (top), peaks resulting from 18-mer emphasized (middle) and peaks resulting from 19-mer emphasized (bottom)

FIGURE 13 is a graphic representation of the process for detecting the Cystic Fibrosis mutation $\Delta F508$ as described in Example 3.

FIGURE 14 is a mass spectrum of the DNA extension product of a $\Delta F508$ homozygous normal of Example 3.

FIGURE 15 is a mass spectrum of the DNA extension product of a $\Delta F508$ heterozygous mutant of Example 3.

-24-

FIGURE 16 is a mass spectrum of the DNA extension product of a Δ F508 homozygous normal of Example 3.

FIGURE 17 is a mass spectrum of the DNA extension product of a Δ F508 homozygous mutant of Example 3.

- 5 FIGURE 18 is a mass spectrum of the DNA extension product of a Δ F508 heterozygous mutant of Example 3.

FIGURE 19 is a graphic representation of various processes for performing apolipoprotein E genotyping of Example 4.

- FIGURE 20 shows the nucleic acid sequence of normal
10 apolipoprotein E (encoded by the E3 allele, FIG. 20B) and other isotypes encoded by the E2 and E4 alleles (FIG. 20A).

FIGURE 21A shows a composite restriction pattern for various genotypes of apolipoprotein E using the CfoI restriction endonuclease.

- FIGURE 21B shows the restriction pattern obtained in a 3.5%
15 MetPhor Agarose Gel for various genotypes of apolipoprotein E.

FIGURE 21C shows the restriction pattern obtained in a 12% polyacrylamide gel for various genotypes of apolipoprotein E.

- FIGURE 22A is a chart showing the molecular weights of the 91, 83, 72, 48 and 35 base pair fragments obtained by restriction enzyme
20 cleavage of the E2, E3 and E4 alleles of apolipoprotein E.

FIGURE 22B is the mass spectrum of the restriction product of a homozygous E4 apolipoprotein E genotype.

FIGURE 23A is the mass spectrum of the restriction product of a homozygous E3 apolipoprotein E genotype.

- 25 FIGURE 23B is the mass spectrum of the restriction product of a E3/E4 apolipoprotein E genotype.

FIGURE 24 is an autoradiograph of Example 5 of a 7.5% polyacrylamide gel in which 10% (5 μ l) of each amplified sample was loaded: sample M: pBR322 A/uI digested; sample 1: HBV positive in

-25-

serological analysis; sample 2: also HBV positive; sample 3: without serological analysis but with an increased level of transaminases, indicating liver disease; sample 4: HBV negative containing HCV; sample 5: HBV positive negative control; (+) positive control). Staining was done
5 with ethidium bromide.

FIGURE 25A is a mass spectrum of sample 1, which is HBV positive. The signal at 20754 Da represents the HBV related amplification product (67 nucleotides, calculated mass: 20735 Da). The mass signal at 10390 Da represents the $[M + 2H]^{2+}$ molecule ion
10 (calculated: 10378 Da).

FIGURE 25B is a mass spectrum of sample 3, which is HBV negative corresponding to nucleic acid (i.e., PCR), serological and dot blot based assays. The amplified product is generated only in trace amounts. Nevertheless it is unambiguously detected at 20751 Da
15 (calculated mass: 20735 Da). The mass signal at 10397 Da represents the $[M + 2H]^{2+}$ molecule ion (calculated: 10376 Da).

FIGURE 25C is a mass spectrum of sample 4, which is HBV negative, but HCV positive. No HBV specific signals were observed.

FIGURE 26 shows a part of the *E. coli lacI* gene with binding sites
20 of the complementary oligonucleotides used in the ligase chain reaction (LCR) of Example 6. Here the wildtype sequence is displayed. The mutant contains a point mutation at bp 191 which is also the site of ligation (bold). The mutation is a C to T transition (G to A, respectively). This leads to a T-G mismatch with oligo B (and A-C mismatch with oligo
25 C, respectively).

FIGURE 27 is a 7.15% polyacrylamide gel of Example 6 stained with ethidium bromide. M: chain length standard (pUC19DNA, *MspI* digested). Lane 1: LCR with wildtype template. Lane 2: LCR with mutant template. Lane 3: (control) LCR without template. The ligation

-26-

product (50 bp) was only generated in the positive reaction containing wildtype template.

FIGURE 28 is an HPLC chromatogram of two pooled positive LCRs.

FIGURE 29 shows an HPLC chromatogram the same conditions but mutant template were used. The small signal of the ligation product is due to either template-free ligation of the educts or to a ligation at a (G-T, A-C) mismatch. The 'false positive' signal is significantly lower than the signal of ligation product with wildtype template depicted in Figure 28. The analysis of ligation educts leads to 'double-peaks' because two of the oligonucleotides are 5'-phosphorylated.

FIGURE 30 In (b) the complex signal pattern obtained by MALDI-TOF-MS analysis of *Pfu* DNA-ligase solution of Example 6 is depicted. In (a) a MALDI-TOF-spectrum of an unpurified LCR is shown. The mass signal 67569 Da probably represents the *Pfu* DNA ligase.

FIGURE 31 shows a MALDI-TOF spectrum of two pooled positive LCRs (a). The signal at 7523 Da represents unligated oligo A (calculated: 7521 Da) whereas the signal at 15449 Da represents the ligation product (calculated: 15450 Da). The signal at 3774 Da is the $[M + 2H]^{2+}$ signal of oligo A. The signals in the mass range lower than 2000 Da are due to the matrix ions. The spectrum corresponds to lane 1 in figure 27 and the chromatogram in figure 28. In (b) a spectrum of two pooled negative LCRs (mutant template) is shown. The signal at 7517 Da represents oligo A (calculated: 7521 Da).

FIGURE 32 shows a spectrum of two pooled control reactions (with salmon sperm DNA as template). The signals in the mass range around 2000 Da are due to Tween20, only oligo A could be detected, as expected.

FIGURE 33 shows a spectrum of two pooled positive LCRs (a). The purification was done with a combination of ultrafiltration and

-27-

streptavidin DynaBeads as described in the text. The signal at 15448 Da represents the ligation product (calculated: 15450 Da). The signal at 7527 represents oligo A (calculated: 7521 Da). The signals at 3761 Da is the $[M + 2H]^{2+}$ signal of oligo A, whereas the signal at 5140 Da is the $[M + 3H]^{2+}$ signal of the ligation product. In (b) a spectrum of two pooled negative LCRs (without template) is shown. The signal at 7514 Da represents oligo A (calculated: 7521 Da).

FIGURE 34 is a schematic presentation of the oligo base extension of the mutation detection primer as described in Example 7, using ddTTP (A) or ddCTP (B) in the reaction mix, respectively. The theoretical mass calculation is given in parenthesis. The sequence shown is part of the exon 10 of the CFTR gene that bears the most common cystic fibrosis mutation $\Delta F508$ and more rare mutations $\Delta I507$ as well as Ile506Ser.

FIGURE 35 is a MALDI-TOF-MS spectrum recorded directly from precipitated oligo base extended primers for mutation detection. The spectrum in (A) and (B), respectively show the annealed primer (CF508) without further extension reaction. Panel C displays the MALDI-TOF spectrum of the wild type by using pppTdd in the extension reaction and D a heterozygotic extension products carrying the 506S mutation when using pppCdd as terminator. Panels E and F show a heterozygote with $\Delta F508$ mutation with pppTdd and pppCdd as terminators in the extension reaction. Panels G and H represent a homozygous $\Delta F508$ mutation with either pppTdd or pppCdd as terminators. The template of diagnosis is pointed out below each spectrum and the observed/expected molecular mass are written in parenthesis.

FIGURE 36 shows the portion of the sequence of pRFc1 DNA, which was used as template for nucleic acid amplification in Example 8 of unmodified and 7-deazapurine containing 99-mer and 200-mer nucleic

-28-

acids as well as the sequences of the 19-mer forward primer and the two 18-mer reverse primers.

FIGURE 37 shows the portion of the nucleotide sequence of M13mp18 RFI DNA, which was used in Example 8 for nucleic acid
5 amplification of unmodified and 7-deazapurine containing 103-mer nucleic acids. Also shown are nucleotide sequences of the 17-mer primers used in the nucleic acid amplification reaction.

FIGURE 38 shows the result of a polyacrylamide gel electrophoresis of amplified products described in Example 8 purified and
10 concentrated for MALDI-TOF MS analysis. M: chain length marker, lane 1: 7-deazapurine containing 99-mer amplified product, lane 2: unmodified 99-mer, lane 3: 7-deazapurine containing 103-mer and lane 4: unmodified 103-mer amplified product.

FIGURE 39: an autoradiogram of polyacrylamide gel
15 electrophoresis of nucleic acid (i.e., PCR) reactions carried out with 5'-[³²P]-labeled primers 1 and 4. Lanes 1 and 2: unmodified and 7-deazapurine modified 103-mer amplified product (53321 and 23520 counts), lanes 3 and 4: unmodified and 7-deazapurine modified 200-mer (71123 and 39582 counts) and lanes 5 and 6: unmodified and 7-
20 deazapurine modified 99-mer (173216 and 94400 counts).

FIGURE 40 a) MALDI-TOF mass spectrum of the unmodified 103-mer amplified products (sum of twelve single shot spectra). The mean value of the masses calculated for the two single strands (31768 u and 31759 u) is 31763 u. Mass resolution: 18. b) MALDI-TOF mass
25 spectrum of 7-deazapurine containing 103-mer amplified product (sum of three single shot spectra). The mean value of the masses calculated for the two single strands (31727 u and 31719 u) is 31723 u. Mass resolution: 67.

-29-

FIGURE 41: a) MALDI-TOF mass spectrum of the unmodified 99-mer amplified product (sum of twenty single shot spectra). Values of the masses calculated for the two single strands: 30261 u and 30794 u. b) MALDI-TOF mass spectrum of 7-deazapurine containing 99-mer amplified product (sum of twelve single shot spectra). Values of the masses calculated for the two single strands: 30224 u and 30750 u.

FIGURE 42: a) MALDI-TOF mass spectrum of the unmodified 200-mer amplified product (sum of 30 single shot spectra). The mean value of the masses calculated for the two single strands (61873 u and 61595 u) is 61734 u. Mass resolution: 28. b) MALDI-TOF mass spectrum of 7-deazapurine containing 200-mer amplified product (sum of 30 single shot spectra). The mean value of the masses calculated for the two single strands (61772 u and 61714 u) is 61643 u. Mass resolution: 39.

FIGURE 43: a) MALDI-TOF mass spectrum of 7-deazapurine containing 100-mer amplified product with ribomodified primers. The mean value of the masses calculated for the two single strands (30529 u and 31095 u) is 30812 u. b) MALDI-TOF mass spectrum of the amplified product after hydrolytic primer-cleavage. The mean value of the masses calculated for the two single strands (25104 u and 25229 u) is 25167 u. The mean value of the cleaved primers (5437 u and 5918 u) is 5677 u.

FIGURE 44 A-D shows the MALDI-TOF mass spectrum of the four sequencing ladders obtained from a 39-mer template (SEQ ID No. 23), which was immobilized to streptavidin beads via a 3' biotinylation. A 14-mer primer (SEQ ID NO. 24) was used in the sequencing according to Example 9.

FIGURE 45 shows a MALDI-TOF mass spectrum of a solid phase sequencing of a 78-mer template (SEQ ID No. 25), which was

-30-

immobilized to streptavidin beads via a 3' biotinylation. A 18-mer primer (SEQ ID No. 26) and ddGTP were used in the sequencing.

FIGURE 46 shows a scheme in which duplex DNA probes with single-stranded overhang capture specific DNA templates and also serve
5 as primers for solid phase sequencing.

FIGURE 47 A-D shows MALDI-TOF mass spectra obtained from a sequencing reaction using 5' fluorescent labeled 23-mer (SEQ ID No. 29) annealed to a 3' biotinylated 18-mer (SEQ ID No. 30), leaving a 5-base overhang, which captured a 15-mer template (SEQ ID No. 31) as
10 described in Example 9.

FIGURE 48 shows a stacking fluorogram of the same products obtained from the reaction described in FIGURE 47, but run on a conventional DNA sequencer.

FIGURE 49 shows a MALDI-TOF mass spectrum of the sequencing
15 ladder using cycle sequencing as described in Example 1 generated from a biological amplified product as template and a 12mer (5'-TGC ACC TGA CTC-3' (SEQ ID NO. 34)) sequencing primer. The peaks resulting from depurinations and peaks which are not related to the sequence are marked by an asterisk. MALDI-TOF MS measurements were taken on a
20 reflectron TOF MS. A.) Sequencing ladder stopped with ddATP; B.) Sequencing ladder stopped with ddCTP; C.) Sequencing ladder stopped with ddGTP; D.) Sequencing ladder stopped with ddTTP.

FIGURE 50 shows a schematic representation of the sequencing
ladder generated in Fig. 49 with the corresponding calculated molecular
25 masses up to 40 bases after the primer. For the calculation, the following masses were used: 3581.4Da for the primer, 312.2 Da for 7-deaza-dATP, 304.2 Da for dTTP, 289.2 Da for dCTP and 328.2 Da for 7-deaza-dGTP.

-31-

FIGURE 51 shows the sequence of the amplified 209 bp amplified product within the β -globin gene, which was used as a template for sequencing. The sequences of the appropriate amplification primer and the location of the 12mer sequencing primer is also shown. This

5 sequence represents a homozygote mutant at the position 4 bases after the primer. In a wildtype sequence this T would be replaced by an A.

FIGURE 52 shows a sequence which is part of the intron 5 of the interferon-receptor gene that bears the AluVpA polymorphism as further described in Example 11. The scheme presents the primer oligo base
10 extension (PROBE) using ddGTP, ddCTP, or both for termination, respectively. The polymorphism detection primer (IFN) is underlined, the termination nucleotides are marked in bold letters. The theoretical mass values from the alleles found in 28 unrelated individuals and a five member family are given in the table. Both second site mutations found
15 in most 13 units allele, but not all, are indicated.

FIGURE 53 shows the MALDI-TOF-MS spectra recorded directly from precipitated extended cyclePROBE reaction products. Family study using AluVpA polymorphism in intron 5 of the interferon- α receptor gene (Example 11).

20 FIGURE 54 shows the mass spectra from PROBE products using ddC as termination nucleotide in the reaction mix. The allele with the molecular mass of approximately 11650 da from the DNA of the mother and child 2 is a hint to a second site mutation within one of the repeat units.

25 FIGURE 55 shows a schematic presentation of the PROBE method for detection of different alleles in the polyT tract at the 3'-end of intron 8 of the CFTR gene with pppCdd as terminator (Example 11).

FIGURE 56 shows the MALDI-TOF-MS spectra recorded directly from the precipitated extended PROBE reaction products. Detection of all

-32-

three common alleles of the polyT tract at the 3' end of Intron 8 of the CFTR gene. (a) T5/T9 heterozygous, (b) T7/T9 heterozygous (Example 11).

FIGURE 57 shows a mass spectrum of the digestion of a 252-mer
5 ApoE gene amplified product ($\epsilon 3/\epsilon 3$ genotype) as described in Example 12 using a) Cfol alone and b) Cfol plus RsaI. Asterisks: depurination peaks.

FIGURE 58 shows a mass spectrum of the ApoE gene amplified
10 product ($\epsilon 3/\epsilon 3$ genotype) digested by Cfol and purified by a) single and b) double ethanol/glycogen and c) double isopropyl alcohol/glycogen precipitations.

FIGURE 59 shows a mass spectrum of the Cfol/RsaI digest
products from a) $\epsilon 2/\epsilon 3$, b) $\epsilon 3/\epsilon 3$, c) $\epsilon 3/\epsilon 4$, and d) $\epsilon 4/\epsilon 4$ genotypes. Dashed lines are drawn through diagnostic fragments.

15 FIGURE 60 shows a scheme for rapid identification of unknown ApoE genotypes following simultaneous digestion of a 252-mer apo E gene amplified product by the restriction enzymes Cfol and RsaI.

FIGURE 61 shows the multiplex (codons 112 and 158) mass
20 spectrum PROBE results for a) $\epsilon 2/\epsilon 3$, b) $\epsilon 3/\epsilon 3$, c) $\epsilon 3/\epsilon 4$, and d) $\epsilon 4/\epsilon 4$ genotypes. E: extension products; P: unextended primer. Top: codon 112 and 158 regions, with polymorphic sites bold and primer sequences underlined.

FIGURE 62 shows a mass spectrum of a TRAP assay to detect
25 telomerase activity (Example 13). The spectrum shows two of the primer signals of the amplified product TS primer at 5,497.3 Da (calc. 5523 Da) and the biotinylated bioCX primer at 7,537.6 Da (calc. 7,537 Da) and the first telomerase-specific assay product containing three telomeric repeats at 12,775.8 Da (calc. 12,452 Da) its mass is larger by

-33-

one dA nucleotide (12,765 Da) due to extendase activity of Taq DNA polymerase.

FIGURE 63 depicts the higher mass range of FIGURE 62, i.e. the peak at 12,775.6 Da represents the products with these telomeric
5 repeats. The peaks at 20,322.1 Da is the result of a telomerase activity to form seven telomeric repeats (calc. 20,395 Da including the extension by one dA nucleotide). The peaks marked 1, 2, 3 and 4 contain a four telomeric repeats at 14,674 Da as well as secondary ion product.

FIGURE 64 displays a MALDI-TOF spectrum of the RT-amplified
10 product of the human tyrosine hydroxylase mRNA indicating the presence of neuroblastoma cells (Example 14). The signal at 18,763.8 Da represents the non-biotinylated single-stranded 61 mer of the nested amplified product (calc. 18,758.2 Da).

FIGURE 65 (a) shows a schematic representation of a PROBE
15 reaction for the RET proto-oncogene with a mixture of dATP, dCTP, dGTP, and ddTTP (Example 15). B represents biotin, through which the sense template strand is bound through streptavidin to a solid support. Figure 65(b) shows the expected PROBE products for ddT and ddA reactions for wildtype, C→T, and C→A antisense strands.

20 FIGURE 66 shows the PROBE product mass spectra for (a) negative control, (b) Patient 1 being heterozygote (Wt/C→T) and (c) Patient 2 being heterozygote (Wt/C→A), reporting average M_r values.

FIGURE 67 shows the MALDI-FTMS spectra for synthetic analogs representing ribo-cleaved RET proto-oncogene amplified products from (a)
25 wildtype, (b) G→A, and (c) G→T homozygotes, and (d) wildtype/G→A, (e) wildtype/G→T, and (f) G→A/G→T heterozygotes, reporting masses of most abundant isotope peaks.

FIGURE 68 is a schematic representation of nucleic acid immobilization via covalent bifunctional trityl linkers.

-34-

FIGURE 69 is a schematic representation of nucleic acid immobilization via hydrophobic trityl linkers.

FIGURE 70 shows a MALDI-TOF mass spectrum of a supernatant of the matrix treated Dynabeads containing bound oligo (5'-iminobiotin -
5 TGCACCTGACTC, SEQ ID NO. 56). An internal standard (CTGTGGTCGTGC, SEQ ID NO. 57) was included in the matrix.

FIGURE 71 shows a MALDI-TOF mass spectrum of a supernatant of the matrix treated Dynabeads containing bound oligo (5'-iminobiotin -
TGCACCTGACTC, SEQ ID NO. 56). An internal standard
10 (CTGTGGTCGTGC, SEQ ID NO. 57) was included in the matrix.

FIGURE 72 schematically depicts the steps involved with the Loop-primer oligo base extension (Loop-probe) reaction.

FIGURE 73A shows a MALDI-TOF mass spectrum of a supernatant after Cfol digest of a stem loop. Figure 73B-D show MALDI-TOF mass
15 spectrum of different genotypes: HbA the wildtype genotype (74B), HbC, a mutation of codon 6 of the β -globin gene which causes sickle cell disease (74C), and HbS, a different mutation of codon 6 of the β -globin gene which causes sickle cell disease (74D).

FIGURE 74 shows the nucleic acid sequence of the amplified
20 region of CKR-5. The underlined sequence corresponds to the region homologous to the amplification primers. The dotted region corresponds to the 32 bp deletion.

FIGURE 75 shows the sense primer ckrT7f. Being designed to facilitate binding of T7-RNA polymerase and amplification of the CKR-5
25 region to be analyzed, it starts with a randomly chosen sequence of 24 bases, the T7 promoter sequence of 18 bases and the sequence homologous to CKR-5 of 19 bases.

-35-

FIGURE 76 is a MALDI-TOF mass spectrum of the CKR-5 amplification product, which was generated as described in the following Example 21.

FIGURE 77 is a positive ion UV-MALDI mass spectra of a synthetic
5 RNA 25-mer (5'-UCCGGUCUGAUGAGUCCGUGAGGAC-3' SEQ ID NO. 62) digested with selected RNAses. For each enzyme 0.6 μ l aliquots of the 4.5 μ l assay containing a total of ca. 20 pmol of the RNA were fixed with 1.5 μ l matrix (3-HPA) for analysis. Fragments with retained 5'-terminus are marked by different arrows, specific for the different
10 RNAses, (Hahner et al., Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, p. 983 (1996)).

FIGURE 78 is an investigation of the specificity of the RNAses CL₃ and Cusativin by positive ion UV-MALDI mass spectra of a synthetic RNA 20-mer. Expected and/or observed cleavage sites are indicated by
15 arrows. A, B, C indicate correct cleavage sites and corresponding singly cleaved fragments. Missing cleavages are designated by a question mark (?), unspecific cleavages by an X.

FIGURE 79 shows the separation of a mixture of DNA molecules (12-mer, 5'-biot. 19-mer, 22-mer and 5'-biot. 27-mer) with streptavidin-coated magnetic beads. a) positive ion UV-MALDI mass spectrum of
20 0.6 μ l of a mixture containing ca. 2-4 pmol of each species mixed with 1.5 μ l matrix (3-HPA). b) same as a) but incubation of the mixture with magnetic beads and subsequent release of the captured fragments.

FIGURE 80 Elution of immobilized 5' biotinylated 49 nt *in vitro*
25 transcript from the streptavidin-coated magnetic beads. Positive UV-MALDI mass spectrum of the transcript prior to incubation with the magnetic beads (a). Spectra of the immobilized RNA transcript after elution with 95% formamide alone (b) and with various additives such as 10mM EDTA (c), 10mM CDTA (d) and 25% ammonium hydroxide (e);

-36-

EDTA and CDTA were adjusted with 25% ammonium hydroxide to a pH of 8.

FIGURE 81 Positive UV-MALDI mass spectra of the 5' biotinylated 49 nt in vitro transcript after RNase U₂ digest for 15 minutes. a)

- 5 Spectrum of the 25 ul assay containing ca. 100 pmol of the target RNA before separation; b) spectrum after isolation of the 5'-biotinylated fragments with magnetic beads. Captured fragments were released by a solution of 95% formamide containing 10 mM CDTA. 1 ul aliquots of the samples were mixed with 1.5 ul matrix (3-HPA) in both cases.

- 10 FIGURE 82 schematically depicts detection of putative mutations in the human β -globin gene at codon 5 and 6 and at codon 30, and the IVS-1 donor site, respectively, done in parallel. FIGURE 82A shows amplification of genomic DNA using the primers β 2 and β 11. The location of the primers and identification tags as well as an indication of
- 15 the wild type and mutant sequences are shown. FIGURE 82B shows analysis of both sites in a simple Primer Reaction Oligo Base Extension (PROBE) using primers β -TAG1 (which binds upstream of codon 5 and 6) and β -TAG2 (which binds upstream of codon 30 and the IVS-1 donor site). Reaction products are captured using streptavidin-coated
- 20 paramagnetic particle bound biotinylated capture primers (cap-tag-1 and cap-tag-2, respectively), that have 6 bases at the 5' end that are complementary to the 5' end of β -TAG1 and β -TAG2, respectively, and a portion which binds to a universal primer.

- FIGURE 83 shows a mass spectrum of the PROBE products of a
- 25 DNA sample from one individual analyzed as described schematically in FIGURE 82.

FIGURE 84 shows a mass spectrum of the sequence bound to cap-tag-2.

-37-

FIGURE 85 shows a mass spectrum obtained by using the β -TAG1 and β -TAG 2 primers in one sequencing reaction using ddATP for termination and then sorting according to the method depicted in FIGURE 82.

- 5 FIGURE 86 shows a mass spectrum obtained by using the β -TAG1 and β -TAG2 primers in one sequencing reaction using ddCTP for termination and then sorting according to the method depicted in FIGURE 82.

- 10 FIGURE 87A shows the wildtype sequence of a fragment of the chemokine receptor CKR-5 gene with primers (bold) used for amplification. The 32 base pair (bp) deletion in the CKR-5 allele is underlined; and the stop nucleotides are in italic. In FIGURE 87B, the wildtype strands are depicted with and without an added Adenosine, their length and molecular masses are indicated. FIGURE 87C indicates
15 the same for the 32 bp deletion. FIGURE 87D shows the PROBE products for the wildtype gene and FIGURE 87E shows the mutated allele.

- 20 FIGURE 88 shows the amplification products of different unrelated individuals as analyzed by native polyacrylamide gel electrophoreses (15%) and silver stain. The band corresponding to a wildtype CKR-5 runs at 75 bp and the band from the gene with the deletion at 43 bp. Bands bigger than 75 bp are due to unspecific amplification.

- 25 FIGURE 89A shows a spectrograph of DNA derived from a heterozygous individual: the peak with a mass of 23319 Da corresponds to the wildtype CKR-5 and the peaks with masses of 13137 Da and 13451 Da to the deletion allele with and without an extra Adenosine, respectively. FIGURE 89B shows a spectrograph of DNA obtained from the same individual as in FIGURE 89A, but the DNA was treated with T4 DNA polymerase to remove the added Adenosine. FIGURES 89C and

-38-

89D are spectrographs derived from homozygous individuals and in FIGURE 89D, the Adenosine has been removed. All peaks with masses lower than 13000 Da are due to multiple charged molecules.

FIGURE 90A shows the mass spectrum of the results of a PROBE
5 reaction performed on DNA obtained from a heterozygous individual. FIGURE 90B shows a mass spectrum of the results of a PROBE reaction on a homozygous individual. The peaks with masses of 6604 Da and 6607 Da, respectively correspond to the wildtype allele, and the peak with a mass of 6275 Da to the deletion allele. The primer is detected
10 with a mass of 5673 and 5676 Da, respectively.

FIGURE 91 shows a MALDI-TOF MS spectra of a thermocycling primer Oligo Base Extension (tc-PROBE) reaction as described in Example 24 using three different templates and 5 different PROBE primers simultaneously in one reaction.

15 FIGURE 92 schematically depicts a single tube process for amplifying and sequencing exons 5-8 of the p53 gene as described in Example 25. The mass spectrum is the A reaction of Figure 93.

FIGURE 93 shows a superposition plot of four separate reactions for sequencing a portion of exon 7 of the p53 gene as described in
20 Example 25.

FIGURE 94 shows the mass spectrum obtained from the A reaction for sequencing a portion of exon 7 of the p53 gene as described in Example 25.

FIGURE 95 shows the mass spectrum of a p53 sequencing ladder
25 for which 5nL of each reaction were transferred to wells of a chip and measured by MALDI-TOF.

FIGURE 96A shows a MALDI-TOF mass spectra of a synthetic 50-mer (15.34 kDa) mixed with 27-mer_{nc} (non-complementary, 8.30 kDa).

-39-

FIGURE 96B shows a MALDI-TOF mass spectra of a synthetic 50-mer (15.34k Da) mixed with a 27-mer_c (complementary, 8.34 kDa). The final concentration of each oligonucleotide was 10 μ M. The signal at 23.68 kDa in Figure 96B corresponds to WC-specific dsDNA.

- 5 FIGURE 97A shows a MALDI-TOF mass spectrum of CfoI/RsaI digest products of a region of exon 4 of the apolipoprotein E gene (ϵ 3 genotype), using sample preparation as in Figure 96.

FIGURE 97B is the same as Figure 97A, except with samples prepared for MALDI-TOF analysis at 4°C.

- 10 FIGURE 98 shows a MALDI-TOF mass spectrum of CfoI/RsaI simultaneously double digest products of a 252 base pair region of exon 4 of the apolipoprotein E gene (ϵ 4 genotype), with samples prepared at 4°C.

- 15 FIGURE 99 shows the mass spectra obtained on a small population study of 15 patients with a 16 element array of diagnostic products transferred to a MALDI target using a pintool microdispenser.

FIGURE 100 is a MALDI mass spectrum of an aliquot sampled after a T₁ digest of a synthetic 20-mer RNA.

DETAILED DESCRIPTION OF THE INVENTION

- 20 **Definitions**

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Where permitted the subject matter of each of the co-pending patent applications and the patent is

- 25 herein incorporated in its entirety.

As used herein, the term "biological sample" refers to any material obtained from any living source (e.g., human, animal, plant, bacteria, fungi, protist, virus). For purposes herein, the biological sample will typically contain a nucleic acid molecule. Examples of appropriate

-40-

biological samples include, but are not limited to: solid materials (e.g., tissue, cell pellets, biopsies) and biological fluids (e.g., urine, blood, saliva, amniotic fluid, mouth wash, cerebral spinal fluid and other body fluids).

- 5 As used herein, the phrases "chain-elongating nucleotides" and "chain-terminating nucleotides" are used in accordance with their art recognized meaning. For example, for DNA, chain-elongating nucleotides include 2'-deoxyribonucleotides (e.g., dATP, dCTP, dGTP and dTTP) and chain-terminating nucleotides include 2', 3'-dideoxyribonucleotides (e.g.,
10 ddATP, ddCTP, ddGTP, ddTTP). For RNA, chain-elongating nucleotides include ribonucleotides (e.g., ATP, CTP, GTP and UTP) and chain-terminating nucleotides include 3'-deoxyribonucleotides (e.g., 3'dA, 3'dC, 3'dG and 3'dU). A complete set of chain elongating nucleotides refers to dATP, dCTP, dGTP and dTTP. The term "nucleotide" is also
15 well known in the art.

- As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified nucleotides such as phosphorothioate nucleotides and deazapurine nucleotides. A complete set of chain-elongating nucleotides refers to four different nucleotides
20 that can hybridize to each of the four different bases comprising the DNA template.

- As used herein, the superscript 0-i designates i + 1 mass differentiated nucleotides, primers or tags. In some instances, the superscript 0 can designate an unmodified species of a particular
25 reactant, and the superscript i can designate the i-th mass-modified species of that reactant. If, for example, more than one species of nucleic acids are to be concurrently detected, then i + 1 different mass-modified detector oligonucleotides (D^0 , D^1 , . . . D^i) can be used to

-41-

distinguish each species of mass modified detector oligonucleotides (D) from the others by mass spectrometry.

As used herein, "multiplexing" refers to the simultaneously detection of more than one analyte, such as more than one (mutated) loci
5 on a particular captured nucleic acid fragment (on one spot of an array).

As used herein, the term "nucleic acid" refers to single-stranded and/or double-stranded polynucleotides such as deoxyribonucleic acid (DNA), and ribonucleic acid (RNA) as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs
10 of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives.

As used herein, the term "conjugated" refers stable attachment, preferably ionic or covalent attachment. Among preferred conjugation means are: streptavidin- or avidin- to biotin interaction; hydrophobic
15 interaction; magnetic interaction (e.g., using functionalized magnetic beads, such as DYNABEADS, which are streptavidin-coated magnetic beads sold by Dynal, Inc. Great Neck, NY and Oslo Norway); polar interactions, such as "wetting" associations between two polar surfaces or between oligo/polyethylene glycol; formation of a covalent bond, such
20 as an amide bond, disulfide bond, thioether bond, or via crosslinking agents; and via an acid-labile or photocleavable linker.

As used herein equivalent, when referring to two sequences of nucleic acids means that the two sequences in question encode the same sequence of amino acids or equivalent proteins. When "equivalent" is
25 used in referring to two proteins or peptides, it means that the two proteins or peptides have substantially the same amino acid sequence with only conservative amino acid substitutions that do not substantially alter the activity or function of the protein or peptide. When "equivalent" refers to a property, the property does not need to be

-42-

present to the same extent [e.g., two peptides can exhibit different rates of the same type of enzymatic activity], but the activities are preferably substantially the same. "Complementary," when referring to two nucleotide sequences, means that the two sequences of nucleotides are

5 capable of hybridizing, preferably with less than 25%, more preferably with less than 15%, even more preferably with less than 5%, most preferably with no mismatches between opposed nucleotides. Preferably the two molecules will hybridize under conditions of high stringency.

As used herein: stringency of hybridization in determining

10 percentage mismatch are those conditions understood by those of skill in the art and typically are substantially equivalent to the following:

- 1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
- 3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

15 It is understood that equivalent stringencies may be achieved using alternative buffers, salts and temperatures.

As used herein, a primer when set forth in the claims refers to a primer suitable for mass spectrometric methods requiring immobilizing, hybridizing, strand displacement, sequencing mass spectrometry refers to

20 a nucleic acid must be of low enough mass, typically about 70 nucleotides or less than 70, and of sufficient size to be useful in the mass spectrometric methods described herein that rely on mass spectrometric detection. These methods include primers for detection and sequencing of nucleic acids, which require a sufficient number

25 nucleotides to form a stable duplex, typically about 6-30, preferably about 10-25, more preferably about 12-20. Thus, for purposes herein a primer will be a sequence of nucleotides comprising about 6-70, more preferably a 12-70, more preferably greater than about 14 to an upper limit of 70, depending upon sequence and application of the primer. The

-43-

primers herein, for example for mutational analyses, are selected to be upstream of loci useful for diagnosis such that when performing using sequencing up to or through the site of interest, the resulting fragment is of a mass that sufficient and not too large to be detected by mass

- 5 spectrometry. For mass spectrometric methods, mass tags or modifier are preferably included at the 5'-end, and the primer is otherwise unlabeled.

As used herein, "conditioning" of a nucleic acid refers to modification of the phosphodiester backbone of the nucleic acid molecule
10 (e.g., cation exchange) for the purpose of eliminating peak broadening due to a heterogeneity in the cations bound per nucleotide unit.

Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide, β -iodoethanol, or 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be

- 15 transformed into a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides that reduce sensitivity for depurination (fragmentation during MS) e.g., a purine analog such as N7- or N9-deazapurine nucleotides, or
20 RNA building blocks or using oligonucleotide triesters or incorporating phosphorothioate functions that are alkylated or employing oligonucleotide mimetics such as peptide nucleic acid (PNA).

As used herein, substrate refers to an insoluble support onto which a sample is deposited according to the materials described herein.

- 25 Examples of appropriate substrates include beads (e.g., silica gel, controlled pore glass, magnetic, agarose gels and crosslinked dextroses (i.e. Sepharose and Sephadex, cellulose and other materials known by those of skill in the art to serve as solid support matrices.

-44-

For examples substrates may be formed from any or combinations of: silica gel, glass, magnet, polystyrene/1% divinylbenzene resins, such as Wang resins, which are Fmoc-amino acid-4-(hydroxymethyl)phenoxymethyl-copoly(styrene-1% divinylbenzene (DVD)) resin, chlorotriyl

- 5 (2-chlorotriylchloride copolystyrene-DVB resin) resin, Merrifield (chloromethylated copolystyrene-DVB) resin metal, plastic, cellulose, cross-linked dextrans, such as those sold under the tradename Sephadex (Pharmacia) and agarose gel, such as gels sold under the tradename Sepharose (Pharmacia), which is a hydrogen bonded polysaccharide-type
- 10 agarose gel, and other such resins and solid phase supports known to those of skill in the art. The support matrices may be in any shape or form, including, but not limited to: capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum, copper and silicon), plastic materials including multiwell plates or
- 15 membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), pins (e.g., arrays of pins suitable for combinatorial synthesis or analysis or beads in pits of flat surfaces such as wafers (e.g., silicon wafers) with or without plates, and beads.

- As used herein, a selectively cleavable linker is a linker that is
- 20 cleaved under selected conditions, such as a photocleavable linker, a chemically cleavable linker and an enzymatically cleavable linker (i.e., a restriction endonuclease site or a ribonucleotide/RNase digestion). The linker is interposed between the support and immobilized DNA.

Isolation of nucleic acids molecules

- 25 Nucleic acid molecules can be isolated from a particular biological sample using any of a number of procedures, which are well-known in the art, the particular isolation procedure chosen being appropriate for the particular biological sample. For example, freeze-thaw and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from

-45-

solid materials; heat and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from urine; and proteinase K extraction can be used to obtain nucleic acid from blood (see, e.g., Rolff et al. (1994) PCR: Clinical Diagnostics and Research, Springer).

- 5 To obtain an appropriate quantity of a nucleic acid molecules on which to perform mass spectrometry, amplification may be necessary. Examples of appropriate amplification procedures for use herein include: cloning (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989), polymerase chain
10 reaction (PCR) (C.R. Newton and A. Graham, PCR, BIOS Publishers, 1994), ligase chain reaction (LCR) (see, e.g., Weidmann et al. (1994) PCR Methods Appl. Vol. 3, Pp. 57-64; F. Barany (1991) Proc. Natl. Acad. Sci. U.S.A. 88:189-93), strand displacement amplification (SDA) (see, e.g., Walker et al. (1994) Nucleic Acids Res. 22:2670-77) and
15 variations such as RT-PCR (see, e.g., Higuchi et al. (1993) Bio/Technology 11:1026-1030), allele-specific amplification (ASA) and transcription based processes.

Immobilization of nucleic acid molecules to solid supports

- To facilitate mass spectrometric analysis, a nucleic acid molecule
20 containing a nucleic acid sequence to be detected can be immobilized to an insoluble (i.e., a solid) support. Examples of appropriate solid supports include beads (e.g., silica gel, controlled pore glass, magnetic, Sephadex/Sepharose, cellulose), capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum,
25 copper and silicon), plastic materials including multiwell plates or membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), pins (e.g., arrays of pins suitable for combinatorial synthesis or analysis or beads in pits of flat surfaces such as wafers (e.g., silicon wafers) with or without filter plates.

-46-

Samples containing target nucleic acids can be transferred to solid supports by any of a variety of methods known to those of skill in the art. For example, nucleic acid samples can be transferred to individual wells of a substrate, e.g., silicon chip, manually or using a pintool
5 microdispenser apparatus as described herein. Alternatively, a piezoelectric pipette apparatus can be used to transfer small nanoliter samples to a substrate permitting the performance of high throughput miniaturized diagnostics on a chip.

Immobilization can be accomplished, for example, based on
10 hybridization between a capture nucleic acid sequence, which has already been immobilized to the support and a complementary nucleic acid sequence, which is also contained within the nucleic acid molecule containing the nucleic acid sequence to be detected (FIGURE 1A). So that hybridization between the complementary nucleic acid molecules is
15 not hindered by the support, the capture nucleic acid can include an e.g., spacer region of at least about five nucleotides in length between the solid support and the capture nucleic acid sequence. The duplex formed will be cleaved under the influence of the laser pulse and desorption can be initiated. The solid support-bound nucleic acid molecule can be
20 presented through natural oligoribo- or oligodeoxyribonucleotide as well as analogs (e.g., thio-modified phosphodiester or phosphotriester backbone) or employing oligonucleotide mimetics such as PNA analogs (see, e.g., Nielsen *et al.*, Science 254:1497 (1991)) which render the base sequence less susceptible to enzymatic degradation and -bound
25 capture base sequence.

-47-

Linkers

A target detection site can be directly linked to a solid support via a reversible or irreversible bond between an appropriate functionality (L') on the target nucleic acid molecule (T) and an appropriate functionality (L) on the capture molecule (FIGURE 1B). A reversible linkage can be such that it is cleaved under the conditions of mass spectrometry (i.e., a photocleavable bond such as a charge transfer complex or a labile bond being formed between relatively stable organic radicals).

- Photocleavable linkers are linkers that are cleaved upon exposure to light (see, e.g., Goldmacher et al. (1992) Bioconj. Chem. 3:104-107), thereby releasing the targeted agent upon exposure to light. Photocleavable linkers that are cleaved upon exposure to light are known (see, e.g., Hazum et al. (1981) in Pept., Proc. Eur. Pept. Symp., 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen et al. (1989) Makromol. Chem 190:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher et al. (1992) Bioconj. Chem. 3:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter et al. (1985) Photochem. Photobiol 42:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. In preferred embodiments, the nucleic acid is immobilized using the photocleavable linker moiety that is cleaved during mass spectrometry. Presently preferred photocleavable linkers are set forth in the EXAMPLES.

-48-

Furthermore, the linkage can be formed with L' being a quaternary ammonium group, in which case, preferably, the surface of the solid support carries negative charges which repel the negatively charged nucleic acid backbone and thus facilitate the desorption required for
5 analysis by a mass spectrometer. Desorption can occur either by the heat created by the laser pulse and/or, depending on L', by specific absorption of laser energy which is in resonance with the L' chromophore.

Thus, the L-L' chemistry can be of a type of disulfide bond
10 (chemically cleavable, for example, by mercaptoethanol or dithioerythrol), a biotin/streptavidin system, a heterobifunctional derivative of a trityl ether group (see, e.g., Köster et al. (1990) "A Versatile Acid-Labile Linker for Modification of Synthetic Biomolecules," Tetrahedron Letters 31:7095) that can be cleaved under mildly acidic conditions as well as
15 under conditions of mass spectrometry, a levulinyl group cleavable under almost neutral conditions with a hydrazinium/acetate buffer, an arginine-arginine or lysine-lysine bond cleavable by an endopeptidase enzyme like trypsin or a pyrophosphate bond cleavable by a pyrophosphatase, or a ribonucleotide bond in between the oligodeoxynucleotide
20 sequence, which can be cleaved, for example, by a ribonuclease or alkali.

The functionalities, L and L', can also form a charge transfer complex and thereby form the temporary L-L' linkage. Since in many cases the "charge-transfer band" can be determined by UV/vis spectrometry (see, e.g., Organic Charge Transfer Complexes by R.
25 Foster, Academic Press, 1969), the laser energy can be tuned to the corresponding energy of the charge-transfer wavelength and, thus, a specific desorption off the solid support can be initiated. Those skilled in the art will recognize that several combinations can serve this purpose

-49-

and that the donor functionality can be either on the solid support or coupled to the nucleic acid molecule to be detected or vice versa.

In yet another approach, a reversible L-L' linkage can be generated by homolytically forming relatively stable radicals. Under the influence of the laser pulse, desorption (as discussed above) as well as ionization will take place at the radical position. Those skilled in the art will recognize that other organic radicals can be selected and that, in relation to the dissociation energies needed to homolytically cleave the bond between them, a corresponding laser wavelength can be selected (see e.g.,
5 Reactive Molecules by C. Wentrup, John Wiley & Sons, 1984).

An anchoring function L' can also be incorporated into a target capturing sequence (TCS) by using appropriate primers during an amplification procedure, such as PCR (FIGURE 4), LCR (FIGURE 5) or transcription amplification (FIGURE 6A).

15 When performing exonuclease sequencing using MALDI-TOF MS, a single stranded DNA molecule immobilized via its 5-end to a solid support is unilaterally degraded with a 3'-processive exonuclease and the molecular weight of the degraded nucleotide is determined sequentially. Reverse Sanger sequencing reveals the nucleotide sequence of the
20 immobilized DNA. By adding a selectively cleavable linker, not only can the mass of the free nucleotides be determined but also, upon removal of the nucleotides by washing, the mass of the remaining fragment can be detected by MALDI-TOF upon cleaving the DNA from the solid support. Using selectively cleavable linkers, such as the photocleavable and
25 chemical cleavable linkers provided herein, this cleavage can be selected to occur during the ionization and volatilizing steps of MALDI-TOF. The same rationale applies for a 5' immobilized strand of a double stranded DNA that is degraded while in a duplex. Likewise, this also

-50-

applies when using a 5'-processive exonuclease and the DNA is immobilized through the 3'-end to the solid support.

As noted, at least three version of immobilization are contemplated herein: 1) the target nucleic acid is amplified or obtained (the target
5 sequence or surrounding DNA sequence must be known to make primers to amplify or isolated); 2) the primer nucleic acid is immobilized to the solid support and the target nucleic acid is hybridized thereto (this is for detecting the presence of or sequencing a target sequence in a sample); or 3) a double stranded DNA (amplified or isolated) is immobilized
10 through linkage to one predetermined strand, the DNA is denatured to eliminate the duplex and then a high concentration of a complementary primer or DNA with identity upstream from the target site is added and a strand displacement occurs and the primer is hybridized to the immobilized strand.

15 In the embodiments where the primer nucleic acid is immobilized on the solid support and the target nucleic acid is hybridized thereto, the inclusion of the cleavable linker allows the primer DNA to be immobilized at the 5'-end so that free 3'-OH is available for nucleic acid synthesis (extension) and the sequence of the "hybridized" target DNA can be
20 determined because the hybridized template can be removed by denaturation and the extended DNA products cleaved from the solid support for MALDI-TOF MS. Similarly for 3), the immobilized DNA strand can be elongated when hybridized to the template and cleaved from the support. Thus, Sanger sequencing and primer oligo base extension
25 (PROBE), discussed below, extension reactions can be performed using an immobilized primer of a known, upstreamn DNA sequence complementary to an invariable region of a target sequence. The nucleic acid from the person is obtained and the DNA sequence of a variable region (deletion, insertion, missense mutation that cause genetic

-51-

predisposition or diseases, or the presence of viral/bacterial or fungal DNA) not only is detected, but the actual sequence and position of the mutation is also determined.

In other cases, the target DNA must be immobilized and the primer
5 annealed. This requires amplifying a larger DNA based on known sequence and then sequencing the immobilized fragments (i.e., the extended fragments are hybridized but not immobilized to the support as described above). In these cases, it is not desirable to include a linker because the MALDI-TOF spectrum is of the hybridized DNA; it is not
10 necessary to cleave the immobilized template.

Any linker known to those of skill in the art for immobilizing nucleic acids to solid supports may be used herein to link the nucleic acid to a solid support. The preferred linkers herein are the selectively cleavable linkers, particularly those exemplified herein. Other linkers
15 include, acid cleavable linkers, such as bismaleimideoxy propane, acid-labile trityl linkers.

Acid cleavable linkers, photocleavable and heat sensitive linkers may also be used, particularly where it may be necessary to cleave the targeted agent to permit it to be more readily accessible to reaction.
20 Acid cleavable linkers include, but are not limited to, bismaleimideoxy propane; and adipic acid dihydrazide linkers (see, e.g., Fattom et al. (1992) Infection & Immun. 60:584-589) and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (see, e.g., Welhöner et
25 al. (1991) J. Biol. Chem. 266:4309-4314).

-52-

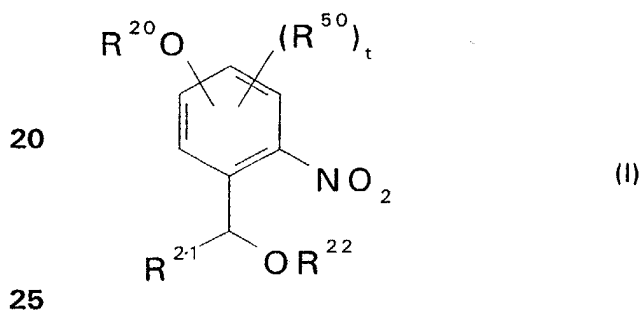
Photocleavable Linkers

Photocleavable linkers are provided. In particular, photocleavable linkers as their phosphoramidite derivatives are provided for use in solid phase synthesis of oligonucleotides. The linkers contain o-nitrobenzyl

5 moieties and phosphate linkages which allow for complete photolytic cleavage of the conjugates within minutes upon UV irradiation. The UV wavelengths used are selected so that the irradiation will not damage the oligonucleotides and are preferably about 350-380 nm, more preferably 365 nm. The photocleavable linkers provided herein possess comparable

10 coupling efficiency as compared to commonly used phosphoramidite monomers (see, Sinha *et al.* (1983) *Tetrahedron Lett.* 24:5843-5846; Sinha *et al.* (1984) *Nucleic Acids Res.* 12:4539-4557; Beaucage *et al.* (1993) *Tetrahedron* 49:6123-6194; and Matteucci *et al.* (1981) *J. Am. Chem. Soc.* 103:3185-3191).

15 In one embodiment, the photocleavable linkers have formula I:

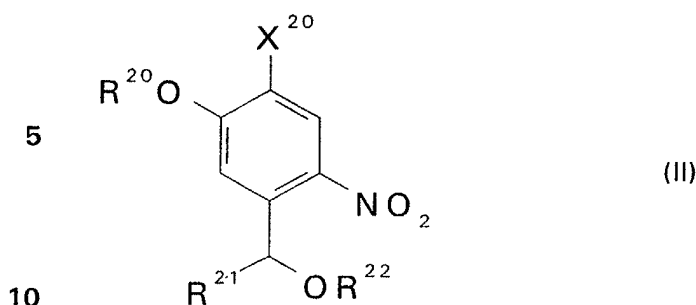


where R²⁰ is ω -(4,4'-dimethoxytrityloxy)alkyl or ω -hydroxyalkyl; R²¹ is selected from hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxy carbonyl and carboxy; R²² is hydrogen or (dialkylamino)(ω -cyanoalkoxy)P-; t is 0-3; and

30 R⁵⁰ is alkyl, alkoxy, aryl or aryloxy.

In a preferred embodiment, the photocleavable linkers have formula II:

-53-



where R^{20} is ω -(4,4'-dimethoxytrityloxy)alkyl, ω -hydroxyalkyl or alkyl; R^{21} is selected from hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxycarbonyl and carboxy; R^{22} is hydrogen or (dialkylamino)(ω -cyanoalkoxy)P-; and X^{20} is hydrogen, alkyl or OR^{20} .

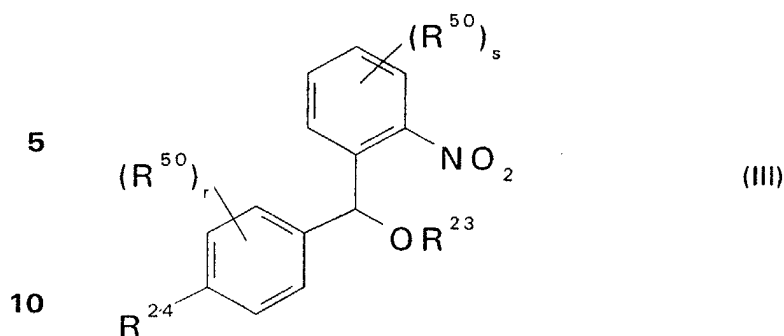
In particularly preferred embodiments, R^{20} is 3-(4,4'-dimethoxytrityloxy)propyl, 3-hydroxypropyl or methyl; R^{21} is selected from hydrogen, methyl and carboxy; R^{22} is hydrogen or (diisopropylamino)(2-cyanoethoxy)P-; and X^{20} is hydrogen, methyl or OR^{20} . In a more preferred embodiment, R^{20} is 3-(4,4'-dimethoxytrityloxy)propyl; R^{21} is methyl; R^{22} is (diisopropylamino)(2-cyanoethoxy)P-; and X^{20} is hydrogen. In another more preferred embodiment, R^{20} is methyl; R^{21} is methyl; R^{22} is (diisopropylamino)(2-cyanoethoxy)P-; and X^{20} is 3-(4,4'-dimethoxytrityloxy)propoxy.

In another embodiment, the photocleavable linkers have formula III:

30

35

-54-



where R^{23} is hydrogen or (dialkylamino)(ω -cyanoalkoxy)P-; and R^{24} is selected from ω -hydroxyalkoxy, ω -(4,4'-dimethoxytrityloxy)alkoxy, ω -hydroxyalkyl and ω -(4,4'-dimethoxytrityloxy)alkyl, and is unsubstituted or substituted on the alkyl or alkoxy chain with one or more alkyl groups; r and s are each independently 0-4; and R^{50} is alkyl, alkoxy, aryl or aryloxy. In certain embodiments, R^{24} is ω -hydroxyalkyl or ω -(4,4'-dimethoxytrityloxy)alkyl, and is substituted on the alkyl chain with a methyl group.

In preferred embodiments, R^{23} is hydrogen or (diisopropylamino)(2-cyanoethoxy)P-; and R^{24} is selected from 3-hydroxypropoxy, 3-(4,4'-dimethoxytrityloxy)propoxy, 4-hydroxybutyl, 3-hydroxy-1-propyl, 1-hydroxy-2-propyl, 3-hydroxy-2-methyl-1-propyl, 2-hydroxyethyl, hydroxymethyl, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)-1-propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-dimethoxytrityloxymethyl.

In more preferred embodiments, R^{23} is (diisopropylamino)(2-cyanoethoxy)P-; r and s are 0; and R^{24} is selected from 3-(4,4'-dimethoxytrityloxy)propoxy, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-

-55-

propyl and 4,4'-dimethoxytrityloxymethyl. R^{24} is most preferably 3-(4,4'-dimethoxytrityloxy)propoxy.

Preparation of the photocleavable linkers

A. Preparation of photocleavable linkers of formulae I or II

5

Photocleavable linkers of formulae I or II may be prepared by the methods described below, by minor modification of the methods by choosing the appropriate starting materials or by any other methods known to those of skill in the art. Detailed procedures for the synthesis of photocleavable linkers of formula II are provided in the Examples.

10

In the photocleavable linkers of formula II where X^{20} is hydrogen, the linkers may be prepared in the following manner. Alkylation of 5-hydroxy-2-nitrobenzaldehyde with an ω -hydroxyalkyl halide, e.g., 3-hydroxypropyl bromide, followed by protection of the resulting alcohol as, e.g., a silyl ether, provides a 5-(ω -silyloxyalkoxy)-2-nitrobenzaldehyde. Addition of an organometallic to the aldehyde affords a benzylic alcohol. Organometallics which may be used include trialkylaluminums (for linkers where R^{21} is alkyl), such as trimethylaluminum, borohydrides (for linkers where R^{21} is hydrogen), such as sodium borohydride, or metal cyanides (for linkers where R^{21} is carboxy or alkoxycarbonyl), such as potassium cyanide. In the case of the metal cyanides, the product of the reaction, a cyanohydrin, would then be hydrolyzed under either acidic or basic conditions in the presence of either water or an alcohol to afford the compounds of interest.

20

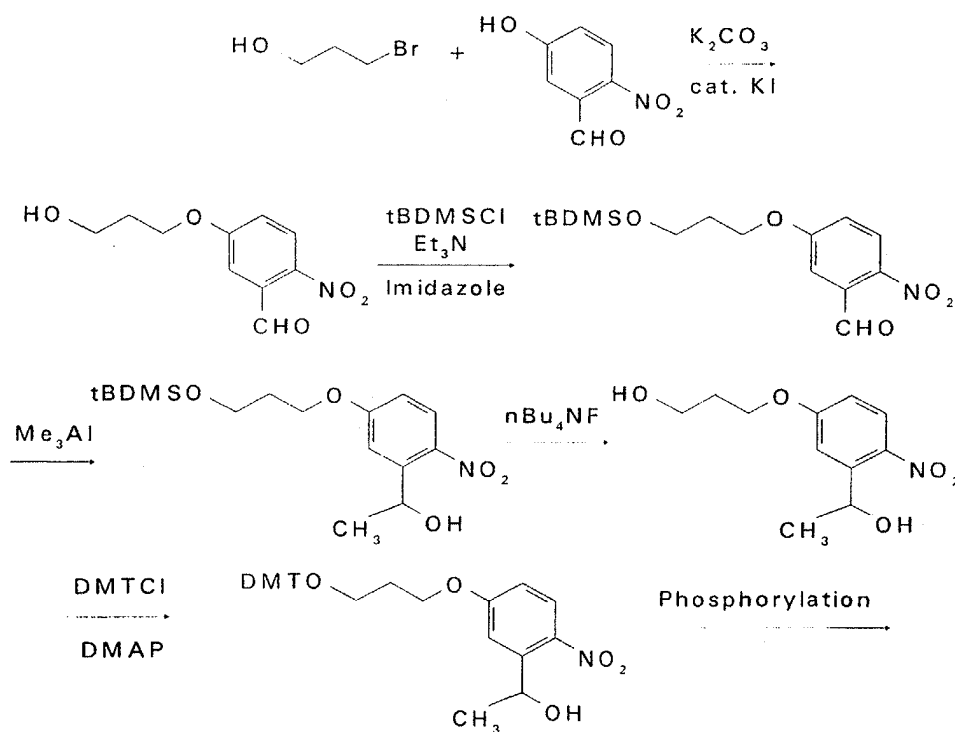
The silyl group of the side chain of the resulting benzylic alcohols may then be exchanged for a 4,4'-dimethoxytriyl group by desilylation with, e.g., tetrabutylammonium fluoride, to give the corresponding alcohol, followed by reaction with 4,4'-dimethoxytrityl chloride. Reaction

25

-56-

with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite affords the linkers where R^{22} is (dialkylamino)(ω -cyanoalkoxy)P-.

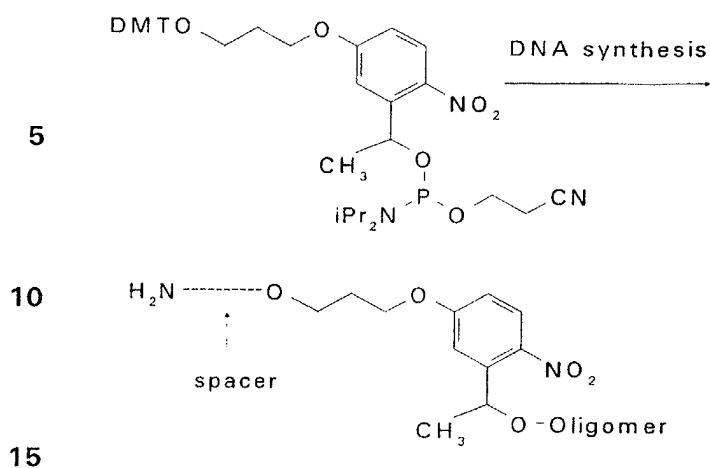
A specific example of a synthesis of a photocleavable linker of formula II is shown in the following scheme, which also demonstrates
 5 use of the linker in oligonucleotide synthesis. This scheme is intended to be illustrative only and in no way limits the scope of the invention. Experimental details of these synthetic transformations are provided in the Examples.



10

15

-57-



Synthesis of the linkers of formula II where X^{20} is OR^{20} , 3,4-dihydroxyacetophenone is protected selectively at the 4-hydroxyl by reaction with, e.g., potassium carbonate and a silyl chloride. Benzoate esters, propiophenones, butyrophenones, etc. may be used in place of the acetophenone. The resulting 4-silyloxy-3-hydroxyacetophenone is then alkylated at the with an alkyl halide (for linkers where R^{20} is alkyl) at the 3-hydroxyl and desilylated with, e.g., tetrabutylammonium fluoride to afford a 3-alkoxy-4-hydroxyacetophenone. This compound is then alkylated at the 4-hydroxyl by reaction with an ω -hydroxyalkyl halide, e.g., 3-hydroxypropyl bromide, to give a 4-(ω -hydroxyalkoxy)-3-alkoxyacetophenone. The side chain alcohol is then protected as an ester, e.g., an acetate. This compound is then nitrated at the 5-position with, e.g., concentrated nitric acid to provide the corresponding 2-nitroacetophenones. Saponification of the side chain ester with, e.g., potassium carbonate, and reduction of the ketone with, e.g., sodium borohydride, in either order gives a 2-nitro-4-(ω -hydroxyalkoxy)-5-alkoxybenzylic alcohol.

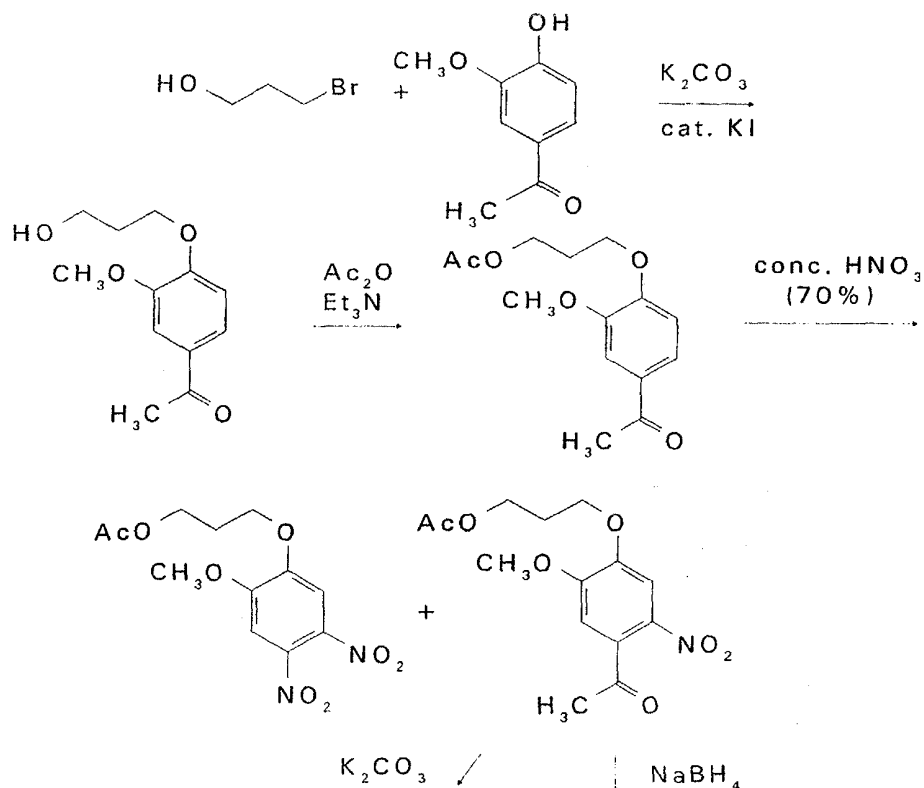
-58-

Selective protection of the side chain alcohol as the corresponding 4,4'-dimethoxytrityl ether is then accomplished by reaction with 4,4'-dimethoxytrityl chloride. Further reaction with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite affords the linkers where R²² is

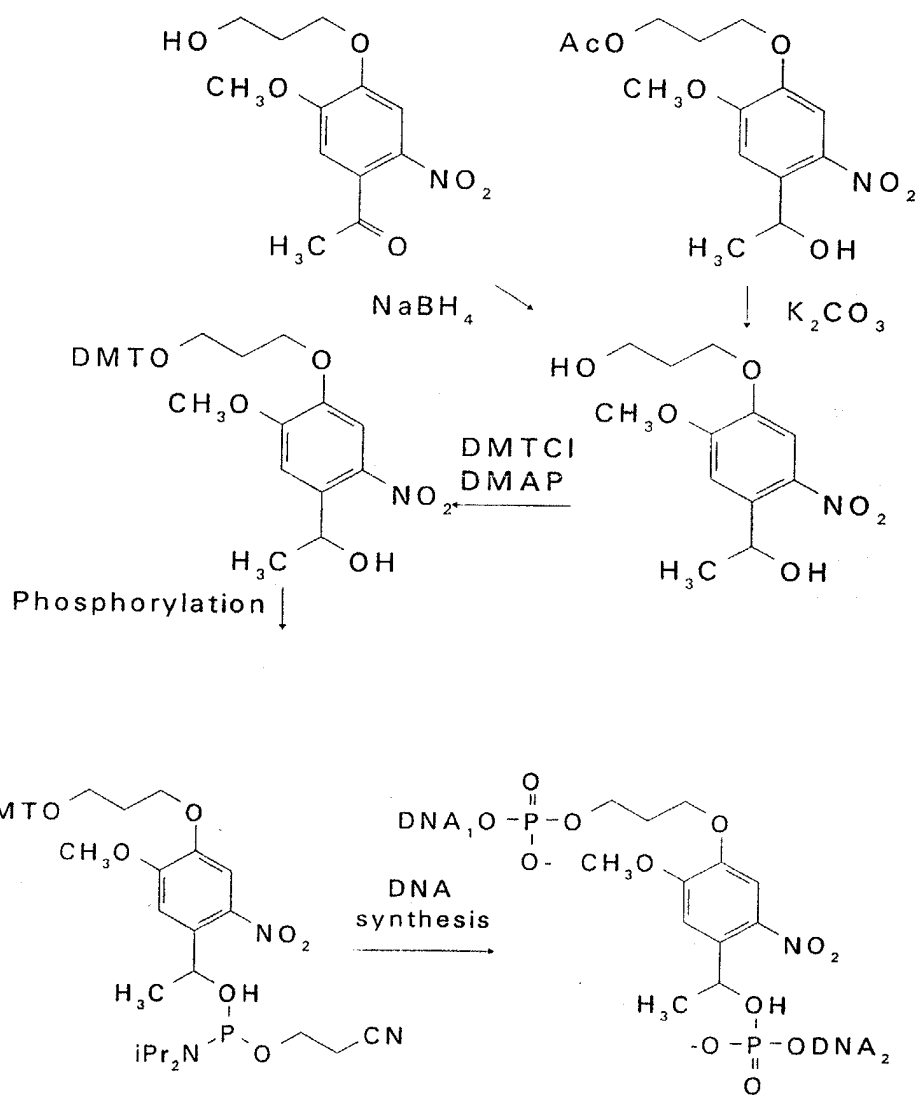
5 (dialkylamino)(ω -cyanoalkoxy)P-.

A specific example of the synthesis of a photocleavable linker of formula II is shown the following scheme. This scheme is intended to be illustrative only and in no way limit the scope of the invention. Detailed experimental procedures for the transformations shown are found in the

10 Examples.



-59-



-60-

B. Preparation of photocleavable linkers of formula III

Photocleavable linkers of formula III may be prepared by the methods described below, by minor modification of the methods by choosing appropriate starting materials, or by other methods known to those of skill in the art.

- In general, photocleavable linkers of formula III are prepared from ω -hydroxyalkyl- or alkoxyaryl compounds, in particular ω -hydroxy-alkyl or alkoxy-benzenes. These compounds are commercially available, or may be prepared from an ω -hydroxyalkyl halide (e.g., 3-hydroxypropyl bromide) and either phenyllithium (for the ω -hydroxyalkylbenzenes) or phenol (for the ω -hydroxyalkoxybenzenes). Acylation of the ω -hydroxyl group (e.g., as an acetate ester) followed by Friedel-Crafts acylation of the aromatic ring with 2-nitrobenzoyl chloride provides a 4-(ω -acetoxy-alkyl or alkoxy)-2-nitrobenzophenone. Reduction of the ketone with, e.g., sodium borohydride, and saponification of the side chain ester are performed in either order to afford a 2-nitrophenyl-4-(hydroxy-alkyl or alkoxy)phenylmethanol. Protection of the terminal hydroxyl group as the corresponding 4,4'-dimethoxytrityl ether is achieved by reaction with 4,4'-dimethoxytrityl chloride. The benzylic hydroxyl group is then reacted with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite to afford linkers of formula II where R²³ is (dialkylamino)(ω -cyanoalkoxy)P-.

- Other photocleavable linkers of formula III may be prepared by substituting 2-phenyl-1-propanol or 2-phenylmethyl-1-propanol for the ω -hydroxy-alkyl or alkoxy-benzenes in the above synthesis. These compounds are commercially available, but may also be prepared by reaction of, e.g., phenylmagnesium bromide or benzylmagnesium bromide, with the requisite oxirane (i.e., propylene oxide) in the presence of catalytic cuprous ion.

-61-

Chemically cleavable linkers

- A variety of chemically cleavable linkers may be used to introduce a cleavable bond between the immobilized nucleic acid and the solid support. Acid-labile linkers are presently preferred chemically cleavable
- 5 linkers for mass spectrometry, especially MALDI-TOF MS, because the acid labile bond is cleaved during conditioning of the nucleic acid upon addition of the 3-HPA matrix solution. The acid labile bond can be introduced as a separate linker group, e.g., the acid labile trityl groups (see Figure 68; Example 16) or may be incorporated in a synthetic
- 10 nucleic acid linker by introducing one or more silyl internucleoside bridges using diisopropylsilyl, thereby forming diisopropylsilyl-linked oligonucleotide analogs. The diisopropylsilyl bridge replaces the phosphodiester bond in the DNA backbone and under mildly acidic conditions, such as 1.5% trifluoroacetic acid (TFA) or 3-HPA/1% TFA
- 15 MALDI-TOF matrix solution, results in the introduction of one or more intra-strand breaks in the DNA molecule. Methods for the preparation of diisopropylsilyl-linked oligonucleotide precursors and analogs are known to those of skill in the art (see e.g., Saha et al. (1993) J. Org. Chem. 58:7827-7831). These oligonucleotide analogs may be readily prepared
- 20 using solid state oligonucleotide synthesis methods using diisopropylsilyl derivatized deoxyribonucleosides.

Nucleic acid conditioning

- Prior to mass spectrometric analysis, it may be useful to "condition" nucleic acid molecules, for example to decrease the laser
- 25 energy required for volatilization and/or to minimize fragmentation. Conditioning is preferably performed while a target detection site is immobilized. An example of conditioning is modification of the phosphodiester backbone of the nucleic acid molecule (e.g., cation exchange), which can be useful for eliminating peak broadening due to a

-62-

- heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide, β -iodoethanol, or 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into
- 5 a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides that reduce sensitivity for depurination (fragmentation during MS) e.g., a purine analog such as N7- or N9-deazapurine nucleotides, or RNA building
- 10 blocks or using oligonucleotide triesters or incorporating phosphorothioate functions which are alkylated or employing oligonucleotide mimetics such as PNA.

Multiplex reactions

- For certain applications, it may be useful to simultaneously detect
- 15 more than one (mutated) loci on a particular captured nucleic acid fragment (on one spot of an array) or it may be useful to perform parallel processing by using oligonucleotide or oligonucleotide mimetic arrays on various solid supports. "Multiplexing" can be achieved by several different methodologies. For example, several mutations can be
- 20 simultaneously detected on one target sequence by employing corresponding detector (probe) molecules (e.g., oligonucleotides or oligonucleotide mimetics). The molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved
- 25 either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide (see FIGURE 2).

-63-

Mass modification of nucleic acids

Mass modifying moieties can be attached, for instance, to either the 5'-end of the oligonucleotide (M^1), to the nucleobase (or bases) (M^2 , M^7), to the phosphate backbone (M^3), and to the 2'-position of the
5 nucleoside (nucleosides) (M^4 , M^6) and/or to the terminal 3'-position (M^5). Examples of mass modifying moieties include, for example, a halogen, an azido, or of the type, XR, wherein X is a linking group and R is a mass-modifying functionality. The mass-modifying functionality can thus be used to introduce defined mass increments into the oligonucleotide
10 molecule.

The mass-modifying functionality can be located at different positions within the nucleotide moiety (see, e.g., U.S. Patent No. 5,547,835 and International PCT application No. WO 94/21822). For example, the mass-modifying moiety, M, can be attached either to the
15 nucleobase, M^2 (in case of the c^7 -dezanucleosides also to C-7, M^7), to the triphosphate group at the alpha phosphate, M^3 , or to the 2'-position of the sugar ring of the nucleoside triphosphate, M^4 and M^6 . Modifications introduced at the phosphodiester bond (M^4), such as with alpha-thio nucleoside triphosphates, have the advantage that these
20 modifications do not interfere with accurate Watson-Crick base-pairing and additionally allow for the one-step post-synthetic site-specific modification of the complete nucleic acid molecule e.g., via alkylation reactions (see, e.g., Nakamaye et al. (1988) Nucl. Acids Res. 16:9947-59). Particularly preferred mass-modifying functionalities are boron-
25 modified nucleic acids since they are better incorporated into nucleic acids by polymerases (see, e.g., Porter et al. (1995) Biochemistry 34:11963-11969; Hasan et al. (1996) Nucleic Acids Res. 24:2150-2157; Li et al. (1995) Nucl. Acids Res. 23:4495-4501).

-64-

Furthermore, the mass-modifying functionality can be added so as to affect chain termination, such as by attaching it to the 3'-position of the sugar ring in the nucleoside triphosphate, M^5 . For those skilled in the art, it is clear that many combinations can be used in the methods

- 5 provided herein. In the same way, those skilled in the art will recognize that chain-elongating nucleoside triphosphates can also be mass-modified in a similar fashion with numerous variations and combinations in functionality and attachment positions.

- Without being bound to any particular theory, the mass-
10 modification, M , can be introduced for X in XR as well as using oligo-/polyethylene glycol derivatives for R . The mass-modifying increment in this case is 44, i.e. five different mass-modified species can be generated by just changing m from 0 to 4 thus adding mass units of 45 ($m=0$), 89 ($m=1$), 133 ($m=2$), 177 ($m=3$) and 221 ($m=4$) to the
15 nucleic acid molecule (e.g., detector oligonucleotide (D) or the nucleoside triphosphates (FIGURE 6(C)), respectively). The oligo/polyethylene glycols can also be monoalkylated by a lower alkyl such as methyl, ethyl, propyl, isopropyl, *t*-butyl and the like. A selection of linking functionalities, X , are also illustrated. Other chemistries can be used in
20 the mass-modified compounds (see, e.g., those described in Oligonucleotides and Analogues, A Practical Approach, F. Eckstein, editor, IRL Press, Oxford, 1991).

- In yet another embodiment, various mass-modifying functionalities, R , other than oligo/polyethylene glycols, can be selected and attached via
25 appropriate linking chemistries, X . A simple mass-modification can be achieved by substituting H for halogens like F , Cl , Br and/or I , or pseudohalogens such as CN , SCN , NCS , or by using different alkyl, aryl or aralkyl moieties such as methyl, ethyl, propyl, isopropyl, *t*-butyl, hexyl, phenyl, substituted phenyl, benzyl, or functional groups such as CH_2F ,

-65-

- CHF₂, CF₃, Si(CH₃)₃, Si(CH₃)₂(C₂H₅), Si(CH₃)(C₂H₅)₂, Si(C₂H₅)₃. Yet another mass-modification can be obtained by attaching homo- or heteropeptides through the nucleic acid molecule (e.g., detector (D)) or nucleoside triphosphates. One example, useful in generating mass-
- 5 modified species with a mass increment of 57, is the attachment of oligoglycines, e.g., mass-modifications of 74 (r=1, m=0), 131 (r=1, m=1), 188 (r=1, m=2), 245 (r=1, m=3) are achieved. Simple oligoamides also can be used, e.g., mass-modifications of 74 (r=1, m=0), 88 (r=2, m=0), 102 (r=3, m=0), 116 (r=4, m=0), etc. are
- 10 obtainable. Variations in additions to those set forth herein will be apparent to the skilled artisan.

- Different mass-modified detector oligonucleotides can be used to simultaneously detect all possible variants/mutants simultaneously (FIGURE 6B). Alternatively, all four base permutations at the site of a
- 15 mutation can be detected by designing and positioning a detector oligonucleotide, so that it serves as a primer for a DNA/RNA polymerase with varying combinations of elongating and terminating nucleoside triphosphates (FIGURE 6C). For example, mass modifications also can be incorporated during the amplification process.

- 20 FIGURE 3 shows a different multiplex detection format, in which differentiation is accomplished by employing different specific capture sequences which are position-specifically immobilized on a flat surface (e.g., a 'chip array'). If different target sequences T1-Tn are present, their target capture sites TCS1-TCSn will specifically interact with
- 25 complementary immobilized capture sequences C1-Cn. Detection is achieved by employing appropriately mass differentiated detector oligonucleotides D1-Dn, which are mass modifying functionalities M1-Mn.

-66-

Mass spectrometric methods for sequencing DNA

Amenable mass spectrometric formats for use herein include the ionization (I) techniques, such as matrix assisted laser desorption ionization (MALDI), electrospray (ESI) (e.g., continuous or pulsed); and
5 related methods (e.g., lonspray, Thermospray, Fast Atomic Bombardment), and massive cluster impact (MCI); these ion sources can be matched with detection formats including lin-linear fields) time-of-flight (TOF), single or multiple quadrupole, single or multiple magnetic sector, Fourier transform ion cyclotron resonance (FTICR), ion trap, or
10 combinations of these to give a hybrid detector (e.g., ion trap - time of flight). For ionization, numerous matrix/wavelength combinations including frozen analyte preparation (MALDI) or solvent combinations (ESI) can be employed.

Since a normal DNA molecule includes four nucleotide units (A, T,
15 C, G), and the mass of each of these is unique (monoisotopic masses 313.06, 304.05, 289.05, 329.05 Da, respectively), an accurate mass determination can define or constrain the possible base compositions of that DNA. Only above 4900 Da does each unit molecular weight have at least one allowable composition; among all 5-mers there is only one non-
20 unique *nominal* molecular weight, among 8-mers, 20. For these and larger oligonucleotides, such mass overlaps can be resolved with the $\sim 1/10^5$ (~ 10 part per million, ppm) mass accuracy available with high resolution FTICR MS. For the 25-mer A_5T_{20} , the 20 composition degeneracies when measured at ± 0.5 Da is reduced to three (A_5T_{20} ,
25 $T_4C_{12}G_9$, $AT_3C_4G_{16}$) when measured with 2 ppm accuracy. Given composition constraints (e.g., the presence or absence of one of the four bases in the strand) can reduce this further (see below).

-67-

Medium resolution instrumentation, including but not exclusively curved field reflectron or delayed extraction time-of-flight MS instruments, can also result in improved DNA detection for sequencing or diagnostics. Either of these are capable of detecting a 9 Da (Δm (A-T)) shift in ≥ 30 -mer strands generated from, for example primer oligo base extension (PROBE), or competitive oligonucleotide single base extension (COSBE), sequencing, or direct detection of small amplified products.

BiomassScan

In this embodiment, exemplified in Example 33, two single stranded nucleic acids are individually immobilized to solid supports. One support contains a nucleic acid encoding the wild type sequence whereas the other support contains a nucleic acid encoding a mutant target sequence. Total human genomic DNA is digested with one or more restriction endonuclease enzyme resulting in the production of small fragments of double stranded genomic DNA (10-1,000 bp). The digested DNA is incubated with the immobilized single stranded nucleic acids and the sample is heated to denature the DNA duplex. The immobilized nucleic acid competes with the other genomic DNA strand for the complementary DNA strand and under the appropriate conditions, a portion of the complementary DNA strand hybridizes to the immobilized nucleic acid resulting in a strand displacement. By using high stringency washing conditions, the two nucleic acids will remain as a DNA duplex only if there is exact identity between the immobilized nucleic acid and the genomic DNA strand. The DNA that remains hybridized to the immobilized nucleic acid is analyzed by mass spectrometry and detection of a signal in the mass spectrum of the appropriate mass is diagnostic for the wild type or mutant allele. In this manner, total genomic DNA can be isolated from a biological sample and screened for the presence or absence of certain mutations. By immobilizing a variety of single

-68-

stranded nucleic acids in an array format, a panel of mutations may be simultaneously screened for a number of genetic loci (i.e., multiplexing).

- In addition, using less stringent washing conditions the hybridized DNA strand may be analyzed by mass spectrometry for changes in the mass resulting from a deletion or insertion within the targeted restriction endonuclease fragment.

Primer oligonucleotide base extension

- As described in detail in the following Example 11, the primer oligo base extension (PROBE) method combined with mass spectrometry identifies the exact number of repeat units (i.e. the number of nucleotides in homogenous stretches) as well as second site mutations within a polymorphic region, which are otherwise only detectable by sequencing. Thus, the PROBE technique increases the total number of detectable alleles at a distinct genomic site, leading to a higher polymorphism information content (PIC) and yielding a far more definitive identification in for instance statistics-based analyses in paternity or forensics applications.

- The method is based on the extension of a detection primer that anneals adjacent to a variable nucleotide tandem repeat (VNTR) or a polymorphic mononucleotide stretch using a DNA polymerase in the presence of a mixture of deoxyNTPs and those dideoxyNTPs that are not present in the deoxy form. The resulting products are evaluated and resolved by MALDI-TOF mass spectrometry without further labeling of the DNA. In a simulated routine application with 28 unrelated individuals, the mass error of this procedure using external calibration was in the worst case 0.38% (56-mer), which is comparable to approximately 0.1 base accuracy; routine standard mass deviations are in the range of 0.1% (.03 bases). Such accuracy with conventional

-69-

electrophoretic methods is not realistic, underscoring the value of PROBE and mass spectrometry in forensic medicine and paternity testing.

The ultra-high resolution of Fourier Transform mass spectrometry makes possible the simultaneous measurement of all reactions of a

- 5 Sanger or Maxam Gilbert sequencing experiment, since the sequence may be read from mass differences instead of base counting from 4 tubes.

- Additionally, the mass differences between adjacent bases generated from unilateral degradation in a stepwise manner by an
- 10 exonuclease can be used to read the entire sequence of fragments generated. Whereas UV or fluorescent measurements will not discriminate mixtures of the nucleoside/nucleotide which are generated when the exonuclease enzyme gets out of phase, this is no problem with mass spectrometry since the resolving power in differentiating between
- 15 the molecular mass of dA, dT, dG and dC is more than significant. The mass of the adjacent bases (i.e., nucleotides) can be determined, for example, using Fast Atomic Bombardment (FAB) or Electrospray Ionization (ESI) mass spectrometry.

- New mutation screening over an entire amplified product can be
- 20 achieved by searching for mass shifted fragments generated in an endonuclease digestion as described in detail in the following Examples 4 and 12.

- Partial sequence information obtained from tandem mass spectrometry (MSⁿ) can place composition constraints as described in the
- 25 preceding paragraph. For the 25-mer above, generation of two fragment ions formed by collisionally activated dissociation (CAD) which differ by 313 Da discounts T₄C₁₂G₉, which contains no A nucleotides; confirming more than a single A eliminates AT₃C₄G₁₆ as a possible composition.

-70-

MSⁿ can also be used to determined full or partial sequences of larger DNAs; this can be used to detect, locate, and identify new mutations in a given gene region. Enzymatic digest products whose masses are correct need not be further analyzed; those with mass shifts

5 could be isolated in real time from the complex mixture in the mass spectrometer and partially sequenced to locate the new mutation.

Table I describes the mutation/polymorphism detection tests that have been developed.

Table I
Mutation/Polymorphism Detection Tests

10

15

20

Clinical Association	Gene	Mutation/Polymorphism
Cystic Fibrosis	CFTR	38 disease causing mutations in 14 exons/introns
Heart Disease (Cholesterol Metabolism)	Apo E Apo A-IV Apo B-100	112R, 112C, 158R, 158C 347S, 347T, 360H, 360Q 3500Q, 3500R
Thyroid Cancer	RET proto-oncogene	C634W, C634T, C634R, C634S, C634F
Sickle Cell Anemia/Thalassemia	beta-globin	Sickle cell anemia S and C 45 thalassemia alleles
HIV Susceptibility	CKR-5	32bp deletion
Breast Cancer Susceptibility	BRCA-2	2bp (AG) deletion in exon 2
Thrombosis	Factor V	R506Q
Arteriosclerosis	GpIIla E-selectin	L33P S128R
Hypertension	ACE	I/D polymorphism

25 Detection of mutations

Diagnosis of genetic diseases

The mass spectrometric processes described above can be used, for example, to diagnose any of the more than 3000 genetic diseases

-71-

currently known (e.g., hemophilias, thalassemias, Duchenne Muscular Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF)) or to be identified.

The following Example 3 provides a mass spectrometric method
5 for detecting a mutation ($\Delta F508$) of the cystic fibrosis transmembrane conductance regulator gene (CFTR), which differs by only three base pairs (900 daltons) from the wild type of CFTR gene. As described further in Example 3, the detection is based on a single-tube, competitive oligonucleotide single base extension (COSBE) reaction using a pair of
10 primers with the 3'-terminal base complementary to either the normal or mutant allele. Upon hybridization and addition of a polymerase and the nucleoside triphosphate one base downstream, only those primers properly annealed (i.e., no 3'-terminal mismatch) are extended; products are resolved by molecular weight shifts as determined by matrix assisted
15 laser desorption ionization time-of-flight mass spectrometry. For the cystic fibrosis $\Delta F508$ polymorphism, 28-mer 'normal' (N) and 30-mer 'mutant' (M) primers generate 29- and 31-mers for N and M homozygotes, respectively, and both for heterozygotes. Since primer and product molecular weights are relatively low (<10 kDa) and the
20 mass difference between these are at least that of a single ~ 300 Da nucleotide unit, low resolution instrumentation is suitable for such measurements.

Thermosequence cycle sequencing, as further described in Example 11, is also useful for detecting a genetic disease.

25 In addition to mutated genes, which result in genetic disease, certain birth defects are the result of chromosomal abnormalities such as Trisomy 21 (Down's Syndrome), Trisomy 13 (Patau Syndrome), Trisomy 18 (Edward's Syndrome), Monosomy X (Turner's Syndrome) and other sex chromosome aneuploidies such as Klienfelter's Syndrome (XXY).

-72-

Here, "house-keeping" genes encoded by the chromosome in question are present in different quantity and the different amount of an amplified fragment compared to the amount in a normal chromosomal configuration can be determined by mass spectrometry.

- 5 Further, there is growing evidence that certain DNA sequences may predispose an individual to any of a number of diseases such as diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancer (e.g., colorectal, breast, ovarian, lung). Also, the detection of "DNA fingerprints", e.g., polymorphisms, such as "mini- and micro-
- 10 satellite sequences", are useful for determining identity or heredity (e.g., paternity or maternity).

The following Examples 4 and 12 provide mass spectrometer based methods for identifying any of the three different isoforms of human apolipoprotein E, which are coded by the E2, E3 and E4 alleles.

- 15 For example, the molecular weights of DNA fragments obtained after restriction with appropriate restriction endonucleases can be used to detect the presence of a mutation and/or a specific allele.

- Depending on the biological sample, the diagnosis for a genetic disease, chromosomal aneuploidy or genetic predisposition can be
- 20 preformed either pre- or post-natally.

Diagnosis of cancer

- Preferred mass spectrometer-based methods for providing an early indication of the existence of a tumor or a cancer are provide herein. For example, as described in Example 13, the telomeric repeat amplification
- 25 protocol (TRAP) in conjunction with telomerase specific extension of a substrate primer and a subsequent amplification of the telomerase specific extension products by an amplification step using a second primer complementary to the repeat structure was used to obtain extension ladders, that were easily detected by MALDI-TOF mass

-73-

spectrometry as an indication of telomerase activity and therefor tumorigenesis.

Alternatively, as described in Example 14, expression of a tumor or cancer associated gene (e.g., human tyrosine 5-hydroxylase) via RT-PCR
5 and analysis of the amplified products by mass spectrometry can be used to detect the tumor or cancer (e.g., biosynthesis of catecholamine via tyrosine 5-hydroxylase is a characteristic of neuroblastoma).

Further, a primer oligo base extension reaction and detection of products by mass spectrometry provides a rapid means for detecting the
10 presence of oncogenes, such as the RET proto oncogene codon 634, which is related to causing multiple endocrine neoplasia, type II (MEN II), as described in Example 15.

Diagnosis of infection

Viruses, bacteria, fungi and other infectious organisms contain
15 distinct nucleic acid sequences, which are different from the sequences contained in the host cell. Detecting or quantitating nucleic acid sequences that are specific to the infectious organism is important for diagnosing or monitoring infection. Examples of disease causing viruses that infect humans and animals and which may be detected by the
20 disclosed processes include: *Retroviridae* (e.g., human immunodeficiency viruses, such as HIV-1 (also referred to as HTLV-III, LAV or HTLV-III/LAV, see, e.g., Ratner et al. (1985) Nature 313: 227-284; Wain-Hobson et al. (1985) Cell 40:9-17); HIV-2 (see, Guyader et al. (1987) Nature 328:662-669 European Patent Publication No. 0 269 520;
25 Chakrabarti et al. (1987) Nature 328:543-547; and European Patent Application No. 0 655 501); and other isolates, such as HIV-LP (International PCT application No. WO 94/00562 entitled "A Novel Human Immunodeficiency Virus"; *Picornaviridae* (e.g., polio viruses, hepatitis A virus, (see, e.g., Gust et al. (1983) Intervirology 20:1-7);

-74-

- entero viruses, human coxsackie viruses, rhinoviruses, echoviruses); *Caliciviridae* (e.g., strains that cause gastroenteritis); *Togaviridae* (e.g., equine encephalitis viruses, rubella viruses); *Flaviridae* (e.g., dengue viruses, encephalitis viruses, yellow fever viruses); *Coronaviridae* (e.g.,
- 5 coronaviruses); *Rhabdoviridae* (e.g., vesicular stomatitis viruses, rabies viruses); *Filoviridae* (e.g., ebola viruses); *Paramyxoviridae* (e.g., parainfluenza viruses, mumps virus, measles virus, respiratory syncytial virus); *Orthomyxoviridae* (e.g., influenza viruses); *Bungaviridae* (e.g., Hantaan viruses, bunga viruses, phleboviruses and Nairo viruses); *Arena*
- 10 *viridae* (hemorrhagic fever viruses); *Reoviridae* (e.g., reoviruses, orbiviruses and rotaviruses); *Birnaviridae*; *Hepadnaviridae* (Hepatitis B virus); *Parvoviridae* (parvoviruses); *Papovaviridae* (papilloma viruses, polyoma viruses); *Adenoviridae* (most adenoviruses); *Herpesviridae* (herpes simplex virus (HSV) 1 and 2, varicella zoster virus,
- 15 cytomegaovirus (CMV), herpes viruses'); *Poxviridae* (variola viruses, vaccinia viruses, pox viruses); and *Iridoviridae* (e.g., African swine fever virus); and unclassified viruses (e.g., the etiological agents of Spongiform encephalopathies, the agent of delta hepatitis (thought to be a defective satellite of hepatitis B virus), the agents of non-A, non-B hepatitis (class
- 20 1 = internally transmitted; class 2 = parenterally transmitted (i.e., Hepatitis C); Norwalk and related viruses, and astroviruses).

Examples of infectious bacteria include, but are not limited to:

- Helicobacter pyloris*, *Borelia burgdorferi*, *Legionella pneumophilia*, *Mycobacteria* sps (e.g., *M. tuberculosis*, *M. avium*, *M. intracellulare*, *M.*
- 25 *kansaii*, *M. gordonae*), *Staphylococcus aureus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Listeria monocytogenes*, *Streptococcus pyogenes* (Group A Streptococcus), *Streptococcus agalactiae* (Group B Streptococcus), *Streptococcus* (viridans group), *Streptococcus faecalis*, *Streptococcus bovis*, *Streptococcus* (anaerobic sps.), *Streptococcus*

-75-

phhenumoniae, pathogenic *Campylobacter* sp., *Enterococcus* sp., *Haemophilus influenzae*, *Bacillus anthracis*, *corynebacterium diphtheriae*, *corynebacterium* sp., *Erysipelothrix rhusiopathiae*, *Clostridium perfringers*, *Clostridium tetani*, *Enterobacter aerogenes*, *Klebsiella*

5 *pneumoniae*, *Pasturella multocida*, *Bacteroides* sp., *Fusobacterium nucleatum*, *Streptobacillus moniliformis*, *Treponema pallidum*, *Treponema pertenu*, *Leptospira*, and *Actinomyces israeli*.

Examples of infectious fungi include: *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Coccidioides immitis*, *Blastomyces dermatitidis*,

10 *Chlamydia trachomatis*, *Candida albicans*. Other infectious organisms (i.e., protists) include: *Plasmodium falciparum* and *Toxoplasma gondii*.

The processes provided herein makes use of the known sequence information of the target sequence and known mutation sites. Although new mutations can also be detected. For example, as shown in FIGURE

15 8, transcription of a nucleic acid molecule obtained from a biological sample can be specifically digested using one or more nucleases and the fragments captured on a solid support carrying the corresponding complementary nucleic acid sequences. Detection of hybridization and the molecular weights of the captured target sequences provide

20 information on whether and where in a gene a mutation is present. Alternatively, DNA can be cleaved by one or more specific endonucleases to form a mixture of fragments. Comparison of the molecular weights between wildtype and mutant fragment mixtures results in mutation detection.

25 **Sequencing by generation of specifically terminated fragementts**

In another embodiment, an accurate sequence determination of a relatively large target nucleic acid, can be obtained by generating specifically terminated fragments from the target nucleic acid, determining the mass of each fragment by mass spectrometry and

-76-

ordering the fragments to determine the sequence of the larger target nucleic acid. In a preferred embodiment, the specifically terminated fragments are partial or complete base-specifically terminated fragments.

One method for generating base specifically terminated fragments

- 5 involves using a base-specific ribonuclease after e.g., a transcription reaction. Preferred base-specific ribonucleases are selected from among: T₁-ribonuclease (G-specific), U₂-ribonuclease (A-specific), PhyM-ribonuclease U specific and ribonuclease A (U/C specific). Other efficient and base-specific ribonucleases can be identified using the assay
- 10 described in Example 16. Preferably modified nucleotides are included in the transcription reaction with unmodified nucleotides. Most preferably, the modified nucleotides and unmodified nucleotides are added to the transcription reaction at appropriate concentrations, so that both moieties are incorporated at a preferential rate of about 1:1. Alternatively, two
- 15 separate transcriptions of the target DNA sequence one with the modified and one with the unmodified nucleotides can be performed and the results compared. Preferred modified nucleotides include: boron or bromine modified nucleotides (Porter et al. (1995) Biochemistry 34:11963-11969; Hasan et al. (1996) Nucl. Acids Res. 24:2150-2157;
- 20 Li et al. (1995) Nucleic Acids Res. 23:4495-4501), α -thio-modified nucleotides, as well as mass-modified nucleotides as described above.

Another method for generating base specifically terminated fragments involves performing a combined amplification and base-specific termination reaction. For example, a combined amplification and

25 termination reaction can be performed using at least two different polymerase enzymes, each having a different affinity for the chain terminating nucleotide, so that polymerization by an enzyme with relatively low affinity for the chain terminating nucleotide leads to exponential amplification whereas an enzyme with relatively high affinity

-77-

for the chain terminating nucleotide terminates the polymerization and yields sequencing products.

- The combined amplification and sequencing can be based on any amplification procedure that employs an enzyme with polynucleotide
- 5 synthetic ability (e.g., polymerase). One preferred process, based on the polymerase chain reaction (PCR), includes the following three thermal steps: 1) denaturing a double stranded (ds) DNA molecule at an appropriate temperature and for an appropriate period of time to obtain the two single stranded (ss) DNA molecules (the template: sense and
- 10 antisense strand); 2) contacting the template with at least one primer that hybridizes to at least one ss DNA template at an appropriate temperature and for an appropriate period of time to obtain a primer containing ss DNA template; 3) contacting the primer containing template at an appropriate temperature and for an appropriate period of time with:
- 15 (i) a complete set of chain elongating nucleotides, (ii) at least one chain terminating nucleotide, (iii) a first DNA polymerase, which has a relatively low affinity towards the chain terminating nucleotide; and (iv) a second DNA polymerase, which has a relatively high affinity towards the chain terminating nucleotide.
- 20 Steps 1)-3) can be sequentially performed for an appropriate number of times (cycles) to obtain the desired amount of amplified sequencing ladders. The quantity of the base specifically terminated fragment desired dictates how many cycles are performed. Although an increased number of cycles results in an increased level of amplification,
- 25 it may also detract from the sensitivity of a subsequent detection. It is therefore generally undesirable to perform more than about 50 cycles, and is more preferable to perform less than about 40 cycles (e.g., about 20-30 cycles).

-78-

Another preferred process for simultaneously amplifying and chain terminating a nucleic acid sequence is based on strand displacement amplification (SDA) (see, e.g., Walker *et al.* (1994) *Nucl. Acids Res.* 22:2670-77; European Patent Publication Number 0 684 315 entitled

- 5 "Strand Displacement Amplification Using Thermophilic Enzymes"). In essence, this process involves the following three steps, which altogether constitute a cycle: 1) denaturing a double stranded (ds) DNA molecule containing the sequence to be amplified at an appropriate temperature and for an appropriate period of time to obtain the two
- 10 single stranded (ss) DNA molecules (the template: sense and antisense strand); 2) contacting the template with at least one primer (P), that contains a recognition/cleavage site for a restriction endonuclease (RE) and that hybridizes to at least one ss DNA template at an appropriate temperature and for an appropriate period of time to obtain a primer
- 15 containing ss DNA template; 3) contacting the primer containing template at an appropriate temperature and for an appropriate period of time with (i) a complete set of chain elongating nucleotides; (ii) at least one chain terminating nucleotide; (iii) a first DNA polymerase, which has a relatively low affinity towards the chain terminating nucleotide; (iv) a second DNA
- 20 polymerase, which has a relatively high affinity towards the chain terminating nucleotide; and (v) an RE that nicks the primer recognition/cleavage site.

- Steps 1)-3) can be sequentially performed for an appropriate number of times (cycles) to obtain the desired amount of amplified
- 25 sequencing ladders. As with the PCR based process, the quantity of the base specifically terminated fragment desired dictates how many cycles are performed. Preferably, less than 50 cycles, more preferably less than about 40 cycles and most preferably about 20 to 30 cycles are performed.

-79-

Preferably about 0.5 to about 3 units of polymerase is used in the combined amplification and chain termination reaction. Most preferably about 1 to 2 units is used. Particularly preferred polymerases for use in conjunction with PCR or other thermal amplification process are

- 5 thermostable polymerases, such as Taq DNA polymerase (Boehringer Mannheim), AmpliTaq FS DNA polymerase (Perkin-Elmer), Deep Vent (exo-), Vent, Vent (exo-) and Deep Vent DNA polymerases (New England Biolabs), Thermo Sequenase (Amersham) or exo(-) *Pseudococcus furiosus* (Pfu) DNA polymerase (Stratagene, Heidelberg, Germany).
- 10 AmpliTaq, Ultman, 9 degree Nm, Tth, Hot Tub, and *Pyrococcus furiosus*. In addition, preferably the polymerase does not have 5'-3' exonuclease activity.

In addition to polymerases, which have a relatively high and a relatively low affinity to the chain terminating nucleotide, a third

- 15 polymerase, which has proofreading capacity (e.g., *Pyrococcus woesei* (Pwo)) DNA polymerase may also be added to the amplification mixture to enhance the fidelity of amplification.

Yet another method for generating base specifically terminated fragments involves contacting an appropriate amount of the target

- 20 nucleic acid with a specific endonuclease or exonuclease. Preferably, the original 5' and/or 3' end of the nucleic acid is tagged to facilitate the ordering of fragments. Tagging of the 3' end is particularly preferred when *in vitro* nucleic acid transcripts are being analyzed, so that the influence of 3' heterogeneity, premature termination and nonspecific
- 25 elongation can be minimized. 5' and 3' tags can be natural (e.g., a 3' poly A tail or 5' or 3' heterogeneity) or artificial. Preferred 5' and/or 3' tags are selected from among the molecules described for mass-modification above.

-80-

The methods provided herein are further illustrated by the following examples, which should not be construed as limiting in any way.

EXAMPLE 1

5 MALDI-TOF desorption of oligonucleotides directly on solid supports

- 1 g CPG (Controlled Pore Glass) was functionalized with 3-(triethoxysilyl)-epoxypropan to form OH-groups on the polymer surface. A standard oligonucleotide synthesis with 13 mg of the OH-CPG on a DNA synthesizer (Milligen, Model 7500) employing β -cyanoethyl-phosphoamidites (Köster *et al.* (1994) *Nucleic Acids Res.* 12:4539) and TAC N-protecting groups (Köster *et al.* (1981) *Tetrahedron* 37:362) was performed to synthesize a 3'-T₅-50mer oligonucleotide sequence in which 50 nucleotides are complementary to a "hypothetical" 50mer sequence. T₅ serves as a spacer. Deprotection with saturated ammonia in methanol at room temperature for 2 hours furnished according to the determination of the DMT group CPG which contained about 10 μ mol 55mer/g CPG. This 55mer served as a template for hybridizations with a 26-mer (with 5'-DMT group) and a 40-mer (without DMT group). The reaction volume is 100 μ l and contains about 1 nmol CPG bound 55mer as template, an equimolar amount of oligonucleotide in solution (26-mer or 40-mer) in 20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂ and 25 mM NaCl. The mixture was heated for 10 min at 65°C and cooled to 37°C during 30' (annealing). The oligonucleotide which has not been hybridized to the polymer-bound template were removed by centrifugation and three subsequent washing/centrifugation steps with 100 μ l each of ice-cold 50 mM ammoniumcitrate. The beads were air-dried and mixed with matrix solution (3-hydroxypicolinic acid/10mM ammonium citrate in acetonitrile/water, 1:1), and analyzed by MALDI-TOF mass spectrometry. The results are presented in Figures 10 and 11.

-81-

EXAMPLE 2**Electrospray (ES) desorption and differentiation of an 18-mer and 19-mer**

DNA fragments at a concentration of 50 pmole/ul in 2-propanol/10mM ammoniumcarbonate (1/9, v/v) were analyzed simultaneously by an electrospray mass spectrometer.

The successful desorption and differentiation of an 18-mer and 19-mer by electrospray mass spectrometry is shown in FIGURE 12.

EXAMPLE 3**10 Detection of The Cystic Fibrosis Mutation Δ F508, by single step dideoxy extension and analysis by MALDI-TOF mass spectrometry (Competitive Oligonucleotide Simple Base Extension = COSBE)**

The principle of the COSBE method is shown in FIGURE 13, N being the normal and M the mutation detection primer, respectively.

MATERIALS AND METHODS

- 15** *PCR Amplification and Strand Immobilization.* Amplification was carried out with exon 10 specific primers using standard PCR conditions (30 cycles: 1'@95°C, 1'@55°C, 2'@72°C); the reverse primer was 5' labelled with biotin and column purified (Oligopurification Cartridge, Cruachem). After amplification the amplified products were purified by
- 20** column separation (Qiagen Quickspin) and immobilized on streptavidin coated magnetic beads (Dynabeads, Dynal, Norway) according to their standard protocol; DNA was denatured using 0.1 M NaOH and washed with 0.1M NaOH, 1xB+W buffer and TE buffer to remove the non-biotinylated sense strand.
- 25** *COSBE Conditions.* The beads containing ligated antisense strand were resuspended in 18 μ l of Reaction mix 1 (2 μ l 10X Taq buffer, 1 μ l (1 unit) Taq Polymerase, 2 μ l of 2 mM dGTP, and 13 μ l H₂O) and incubated at 80°C for 5' before the addition of Reaction mix 2 (100 ng each of COSBE primers). The temperature was reduced to 60°C and the
- 30** mixtures incubated for a 5' annealing/extension period; the beads were

-82-

then washed in 25mM triethylammonium acetate (TEAA) followed by 50mM ammonium citrate.

- Primer Sequences.* All primers were synthesized on a Perseptive Biosystems Expedite 8900 DNA Synthesizer using conventional
- 5 phosphoramidite chemistry (Sinha *et al.* (1984) Nucleic Acids Res. 12:4539). COSBE primers (each containing an intentional mismatch one base before the 3'-terminus) were those used in a previous ARMS study (Ferrie *et al.* (1992) Am J Hum Genet 51:251-262) with the exception that two bases were removed from the 5'-end of the normal:
- 10 Ex10 PCR (Forward): 5'-BIO-GCA AGT GAA TCC TGA GCG TG-3'
(SEQ ID No. 1)
Ex10 PCR (Reverse): 5'-GTG TGA AGG GTT CAT ATG C-3'
(SEQ ID No. 2)
COSBE ΔF508-N 5'-ATC TAT ATT CAT CAT AGG AAA CAC CAC A-3'
15 (28-mer) (SEQ ID No. 3)
COSBE ΔF508-N 5'-GTA TCT ATA TTC ATC ATA GGA AAC ACC ATT-
3' (30-mer) (SEQ ID No. 4)

- Mass Spectrometry.* After washing, beads were resuspended in 1 μL 18 Mohm/cm H₂O. 300 nL each of matrix (Wu *et al.* (1993) Rapid
- 20 Commun. Mass Spectrom. 7:142-146) solution (0.7 M 3-hydroxypicolinic acid, 0.7 M dibasic ammonium citrate in 1:1 H₂O:CH₃CN) and resuspended beads (Tang *et al.* (1995) Rapid Commun Mass Spectrom 8:727-730) were mixed on a sample target and allowed to air dry. Up to 20 samples were spotted on a probe target disk for introduction into the
- 25 source region of an unmodified Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular weights (M_r(calc)) were calculated from atomic compositions. Vendor provided software was used to determine peak centroids using

-83-

external calibration; 1.08 Da has been subtracted from these to correct for the charge carrying proton mass to yield the text $M_r(\text{exp})$ values.

Scheme. Upon annealing to the bound template, the N and M primers (8508.6 and 9148.0 Da, respectively) are presented with dGTP; only primers with proper Watson-Crick base paring at the variable (V) position are extended by the polymerase. Thus if V pairs with the 3'-terminal base of N, N is extended to a 8837.9 Da product (N + 1).

Likewise, if V is properly matched to the M terminus, M is extended to a 9477.3 Da M + 1 product.

10 *Results*

Figures 14 - 18 show the representative mass spectra of COSBE reaction products. Better results were obtained when amplified products were purified before the biotinylated anti-sense strand was bound.

EXAMPLE 4

15 **Differentiation of Human Apolipoprotein E Isoforms by Mass Spectrometry**

Apolipoprotein E (Apo E), a protein component of lipoproteins, plays an essential role in lipid metabolism. For example, it is involved with cholesterol transport, metabolism of lipoprotein particles,

20 immunoregulation and activation of a number of lipolytic enzymes.

There are three common isoforms of human Apo E (coded by E2, E3 and E4 alleles). The most common is the E3 allele. The E2 allele has been shown to decrease the cholesterol level in plasma and therefore may have a protective effect against the development of atherosclerosis.

25 The DNA encoding a portion of the E2 allele is set forth in SEQ ID No. 130. Finally, the E4 isoform has been correlated with increased levels of cholesterol, conferring predisposition to atherosclerosis. Therefore, the identity of the apo E allele of a particular individual is an important determinant of risk for the development of cardiovascular disease.

-84-

As shown in Figure 19, a sample of DNA encoding apolipoprotein E can be obtained from a subject, amplified (e.g., via PCR); and the amplified product can be digested using an appropriate enzyme (e.g., CfoI). The restriction digest obtained can then be analyzed by a variety
5 of means. As shown in Figure 20, the three isotypes of apolipoprotein E (E2, E3 and E4 have different nucleic acid sequences and therefore also have distinguishable molecular weight values.

As shown in Figure 21A-C, different Apolipoprotein E genotypes exhibit different restriction patterns in a 3.5% MetPhor Agarose Gel or
10 12% polyacrylamide gel. As shown in Figures 22 and 23, the various apolipoprotein E genotypes can also be accurately and rapidly determined by mass spectrometry.

EXAMPLE 5

Detection of hepatitis B virus in serum samples.

15 MATERIALS AND METHODS

Sample preparation

Phenol/choloform extraction of viral DNA and the final ethanol precipitation was done according to standard protocols.

First PCR

20 Each reaction was performed with 5 μ l of the DNA preparation from serum. 15 pmol of each primer and 2 units Taq DNA polymerase (Perkin Elmer, Weiterstadt, Germany) were used. The final concentration of each dNTP was 200 μ MM, the final volume of the reaction was 50 μ l. 10x PCR buffer (Perkin Elmer, Weiterstadt, Germany) contained 100 mM
25 Tris-HCl, pH 8.3, 500 mM KCl, 15 mM MgCl₂, 0.01% gelatine (w/v).
Primer sequences:

Primer	SEQUENCE	SEQ ID No.
1	5'-GCTTTGGGCGCATGGACATTGACCCGTATAA-3'	5
2	5'-CTGACTACTAATTCCTGGATGCTGGGTCT-3'	6

-85-

Nested PCR:

Each reaction was performed either with 1 μ l of the first reaction or with a 1:10 dilution of the first PCR as template, respectively. 100 pmol of each primer, 2.5 u *Pfu*(exo-) DNA polymerase (Stratagene, Heidelberg, Germany), a final concentration of 200 μ M of each dNTPs and 5 μ l 10x *Pfu* buffer (200 mM Tris-HCl, pH 8.75, 100 mM KCl, 100 mM (NH₄)₂SO₄, 1% Triton X-100, 1 mg/ml BSA, (Stratagene, Heidelberg, Germany) were used in a final volume 50 μ l. The reactions were performed in a thermocycler (OmniGene, MWG-Biotech, Ebersberg, Germany) using the following program: 92°C for 1 minute, 60°C for 1 minute and 72°C for 1 minute with 20 cycles. Sequence of oligodeoxynucleotides (purchased HPLC-purified from MWG-Biotech, Ebersberg, Germany):

- HBV13: 5'-TTGCCTGAGTGCAGTATGGT-3' (SEQ ID NO. 7)
15 HBV15bio: Biotin-5'-AGCTCTATATCGGGAAGCCT-3' (SEQ ID NO. 8)

Purification of amplified products:

For the recording of each spectrum, one PCR, 50 μ l, (performed as described above) was used. Purification was done according to the following procedure: Ultrafiltration was done using Ultrafree-MC filtration units (Millipore, Eschborn, Germany) according to the protocol of the provider with centrifugation at 8000 rpm for 20 minutes. 25 μ l (10 μ g/ μ l) streptavidin Dynabeads (Dyna, Hamburg, Germany) were prepared according to the instructions of the manufacturer and resuspended in 25 μ l of B/W buffer (10 mM Tris-HCl, pH 7.5, 1mM EDTA, 2 M NaCl).
25 This suspension was added to the PCR samples still in the filtration unit and the mixture was incubated with gentle shaking for 15 minutes at ambient temperature. The suspension was transferred in a 1.5 ml Eppendorf tube and the supernatant was removed with the aid of a Magnetic Particle Collector, MPC, (Dyna, Hamburg, Germany). The

-86-

beads were washed twice with 50 μ l of 0.7 M ammonium citrate solution, pH 8.0 (the supernatant was removed each time using the MPC). Cleavage from the beads can be accomplished by using formamide at 90°C. The supernatant was dried in a speedvac for about
5 an hour and resuspended in 4 μ l of ultrapure water (MilliQ UF plus Millipore, Eschborn, Germany). This preparation was used for MALDI-TOF MS analysis.

MALDI-TOF MS:

Half a microliter of the sample was pipetted onto the sample
10 holder, then immediately mixed with 0.5 μ l matrix solution (0.7 M 3-hydroxypicolinic acid 50% acetonitrile, 70 mM ammonium citrate). This mixture was dried at ambient temperature and introduced into the mass spectrometer. All spectra were taken in positive ion mode using a Finnigan MAT Vision 2000 (Finnigan MAT, Bremen, Germany), equipped
15 with a reflectron (5 keV ion source, 20 keV postacceleration) and a 337 nm nitrogen laser. Calibration was done with a mixture of a 40-mer and a 100-mer. Each sample was measured with different laser energies. In the negative samples, the amplified product was detected neither with less nor with higher laser energies. In the positive samples the amplified
20 product was detected at different places of the sample spot and also with varying laser energies.

RESULTS

A nested PCR system was used for the detection of HBV DNA in blood samples employing oligonucleotides complementary to the c region
25 of the HBV genome (primer 1: beginning at map position 1763, primer 2 beginning at map position 2032 of the complementary strand) encoding the HBV core antigen (HBVcAg). DNA was isolated from patients serum according to standard protocols. A first PCR was performed with the

-87-

DNA from these preparations using a first set of primers. If HBV DNA was present in the sample a DNA fragment of 269 bp was generated.

- In the second reaction, primers which were complementary to a region within the PCR fragment generated in the first PCR were used. If
- 5 HBV related amplified products were present in the first PCR a DNA fragment of 67 bp was generated (see Fig. 25A) in this nested PCR. The usage of a nested PCR system for detection provides a high sensitivity and also serves as a specificity control for the external PCR (Rolfs *et al.* (1992) PCR: Clinical Diagnostics and Research, Springer, Heidelberg). A
- 10 further advantage is that the amount of fragments generated in the second PCR is high enough to ensure an unproblematic detection although purification losses can not be avoided.

- The samples were purified using ultrafiltration to restreptavidin Dynabeads. This purification was done because the shorter primer
- 15 fragments were immobilized in higher yield on the beads due to stearic reasons. The immobilization was done directly on the ultrafiltration membrane to avoid substance losses due to unspecific absorption on the membrane. Following immobilization, the beads were washed with ammonium citrate to perform cation exchange (Pieles *et al.* (1993) Nucl.
- 20 Acids Res. 21:3191-3196). The immobilized DNA was cleaved from the beads using 25% ammonia which allows cleavage of DNA from the beads in a very short time, but does not result in an introduction of sodium or other cations.

- The nested PCRs and the MALDI TOF analysis were performed
- 25 without knowing the results of serological analysis. Due to the unknown virus titer, each sample of the first PCR was used undiluted as template and in a 1:10 dilution, respectively.

Sample 1 was collected from a patient with chronic active HBV infection who was positive in Hbs- and Hbe-antigen tests but negative in

-88-

a dot blot analysis. Sample 2 was a serum sample from a patient with an active HBV infection and a massive viremia who was HBV positive in a dot blot analysis. Sample 3 was a denatured serum sample therefore no serological analysis could be performed by an increased level of transaminases indicating liver disease was detected. In autoradiograph analysis (Figure 24), the first PCR of this sample was negative. Nevertheless, there was some evidence of HBV infection. This sample is of interest for MALDI-TOF analysis, because it demonstrates that even low-level amounts of amplified products can be detected after the purification procedure. Sample 4 was from a patient who was cured of HBV infection. Samples 5 and 6 were collected from patients with a chronic active HBV infection.

Figure 24 shows the results of a PAGE analysis of the nested PCR reaction. A amplified product is clearly revealed in samples 1, 2, 3, 5 and 6. In sample 4 no amplified product was generated, it is indeed HBV negative, according to the serological analysis. Negative and positive controls are indicated by + and -, respectively. Amplification artifacts are visible in lanes 2, 5, 6 and + if non-diluted template was used. These artifacts were not generated if the template was used in a 1:10 dilution. In sample 3, amplified product was merely detectable if the template was not diluted. The results of PAGE analysis are in agreement with the data obtained by serological analysis except for sample 3 as discussed above.

Figure 25A shows a mass spectrum of a nested amplified product from sample number 1 generated and purified as described above. The signal at 20754 Da represents the single stranded amplified product (calculated: 20735 Da, as the average mass of both strands of the amplified product cleaved from the beads). The mass difference of calculated and obtained mass is 19 Da (0.09%). As shown in Fig. 25A,

-89-

sample number 1 generated a high amount of amplified product, resulting in an unambiguous detection.

Fig. 25B shows a spectrum obtained from sample number 3. As depicted in Fig. 24, the amount of amplified product generated in this section is significantly lower than that from sample number 1.

Nevertheless, the amplified product is clearly revealed with a mass of 20751 Da (calculated 20735). The mass difference is 16 Da (0.08%).

The spectrum depicted in Fig. 25C was obtained from sample number 4 which is HBV negative (as is also shown in Fig 24). As expected no

signals corresponding to the amplified product could be detected. All samples shown in Fig. 25 were analyzed with MALDI-TOF MS, whereby amplified product was detected in all HBV positive samples, but not in the HBV negative samples. These results were reproduced in several independent experiments.

EXAMPLE 6

Analysis of Ligase Chain Reaction Products Via MALDI-TOF Mass Spectrometry

MATERIALS AND METHODS

Oligodeoxynucleotides

Except the biotinylated one and all other oligonucleotides were synthesized in a 0.2 μ mol scale on a MilliGen 7500 DNA Synthesizer (Millipore, Bedford, MA, USA) using the β -cyanoethylphosphoramidite method (Sinha, N.D. et al. (1984) Nucleic Acids Res. 12:4539-4577).

The oligodeoxynucleotides were RP-HPLC-purified and deprotected

according to standard protocols. The biotinylated oligodeoxynucleotide was purchased (HPLC-purified) from Biometra, Gottingen, Germany).

Sequences and calculated masses of the oligonucleotides used:

Oligodeoxy-nucleotide	SEQUENCE	SEQ ID No.
A	5'-p-TTGTGCCACGCGGTTGGGAATGTA (7521 Da)	9

-90-

B	5'-p-AGCAACGACTGTTTGCCCGCCAGTTG (7948 Da)	10
C	5'-bio-TACATTCCCAACCGCGTGGCACAAC (7960 Da)	11
D	5'-p-AACTGGCGGGCAAACAGTCGTTGCT (7708 Da)	12

5 *5-Phosphorylation of oligonucleotides A and D*

This was performed with polynucleotide kinase (Boehringer, Mannheim, Germany) according to published procedures, the 5'-phosphorylated oligonucleotides were used unpurified for LCR.

Ligase chain reaction

10 The LCR was performed with *Pfu* DNA ligase and a ligase chain reaction kit (Stratagene, Heidelberg, Germany) containing two different pBluescript KII phagemids. One carrying the wildtype form of the *E.coli lacI* gene and the other one a mutant of this gene with a single point mutation at bp 191 of the *lacI* gene.

15 The following LCR conditions were used for each reaction: 100 pg template DNA (0.74 fmol) with 500 pg sonified salmon sperm DNA as carrier, 25 ng (3.3 pmol) of each 5'-phosphorylated oligonucleotide, 20 ng (2.5 pmol) of each non-phosphorylated oligonucleotide, 4 U *Pfu* DNA ligase in a final volume of 20 μ l buffered ss 50-mer was used (1 fmol) as
 20 template, in this case oligo C was also biotinylated. All reactions were performed in a thermocycler (OmniGene, MWG-Biotech, Ebersberg, Germany) with the following program: 4 minutes 92°C, 2 minutes 60°C and 25 cycles of 20 seconds 92°C, 40 seconds 60°C. Except for HPLC analysis the biotinylated ligation educt C was used. In a control
 25 experiment the biotinylated and non-biotinylated oligonucleotides revealed the same gel electrophoretic results. The reactions were analyzed on 7.5% polyacrylamide gels. Ligation product 1 (oligo A and B) calculated mass: 15450 Da, ligation product 2 (oligo C and D) calculated mass: 15387 Da.

30 *SMART-HPLC*

-91-

Ion exchange HPLC (IE HPLC) was performed on the SMART-system (Pharmacia, Freiburg, Germany) using a Pharmacia Mono Q, PC 1.6/5 column. Eluents were buffer A (25 mM Tris-HCl, 1 mM EDTA and 0.3 M NaCl at pH 8.0) and buffer B (same as A, but 1 M NaCl). Starting with 100% A for 5 minutes at a flow rate of 50 μ l/min. a gradient was applied from 0 to 70% B in 30 minutes, then increased to 100% B in 2 minutes and held at 100% B for 5 minutes. Two pooled LCR volumes (40 μ l) performed with either wildtype or mutant template were injected.

10 *Sample preparation for MALDI-TOF-MS*

Preparation of immobilized DNA: For the recording of each spectrum two LCRs (performed as described above) were pooled and diluted 1:1 with 2x B/W buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl). To the samples 5 μ l streptavidin DynaBeads (Dyna, Hamburg, Germany) were added, the mixture was allowed to bind with gentle shaking for 15 minutes at ambient temperature. The supernatant was removed using a Magnetic Particle Collector, MPC, (Dyna, Hamburg, Germany) and the beads were washed twice with 50 μ l of 0.7 M ammonium citrate solution (pH 8.0) (the supernatant was removed each time using the MPC). The beads were resuspended in 1 μ l of ultrapure water (MilliQ, Millipore, Bedford, Mabelow).

Combination of ultrafiltration and streptavidin DynaBeads: For the recording of spectrum two LCRs (performed as described above) were pooled, diluted 1:1 with 2x B/W buffer and concentrated with a 5000 NMWL Ultrafree-MC filter unit (Millipore, Eschborn, Germany) according to the instructions of the manufacturer. After concentration the samples were washed with 300 μ l 1x B/W buffer to streptavidin DynaBeads were added. The beads were washed once on the Ultrafree-MC filtration unit with 300 μ l of 1x B/W buffer and processed as described above. The

-92-

beads were resuspended in 30 to 50 μ l of 1x B/W buffer and transferred in a 1.5 ml Eppendorf tube. The supernatant was removed and the beads were washed twice with 50 μ l of 0.7 M ammonium citrate (pH 8.0). Finally, the beads were washed once with 30 μ l of acetone and
5 resuspended in 1 μ l of ultrapure water. The ligation mixture after immobilization on the beads was used for MALDS-TOF-MS analysis as described below.

MALDI-TOF-MS

A suspension of streptavidin-coated magnetic beads with the
10 immobilized DNA was pipetted onto the sample holder, then immediately mixed with 0.5 μ l matrix solution (0.7 M 3-hydroxypicolinic acid in 50% acetonitrile, 70 mM ammonium citrate). This mixture was dried at ambient temperature and introduced into the mass spectrometer. All spectra were taken in positive ion mode using a Finnigan MAT Vision
15 2000 (Finnigan MAT, Bremen, Germany), equipped with a reflectron (5 keV ion source, 20 keV postacceleration) and a nitrogen laser (337 nm). For the analysis of *Pfu* DNA ligase 0.5 μ l of the solution was mixed on the sample holder with 1 μ l of matrix solution and prepared as described above. For the analysis of unpurified LCRs 1 μ l of an LCR was mixed
20 with 1 μ l matrix solution.

RESULTS

The *E. coli lacI* gene served as a simple model system to investigate the suitability of MALDI-TOF-MS as detection method for products generated in ligase chain reactions. This template system
25 contains of an *E. coli lacI* wildtype gene in a pBluescript KII phagemid and an *E. coli lacI* gene carrying a single point mutation at bp 191 (C to T transition; SEQ ID No. 131) in the same phagemid. Four different oligonucleotides were used, which were ligated only if the *E coli lacI* wildtype gene was present (Figure 26).

-93-

LCR conditions were optimized using *Pfu* DNA ligase to obtain at least 1 pmol ligation product in each positive reaction. The ligation reactions were analyzed by polyacrylamide gel electrophoresis (PAGE) and HPLC on the SMART system (Figures 27, 28 and 29). Figure 27 shows a PAGE of a positive LCR with wildtype template (lane 1), a negative LCR with mutant template (1 and 2) and a negative control which contains enzyme, oligonucleotides and no template but salmon sperm DNA. The gel electrophoresis clearly shows that the ligation product (50 bp) was produced only in the reaction with wildtype template; whereas neither the template carrying the point mutation nor the control reaction with salmon sperm DNA generated amplification products. In Figure 28, HPLC was used to analyze two pooled LCRs with wildtype template performed under the same conditions. The ligation product was clearly revealed. Figure 29 shows the results of a HPLC in which two pooled negative LCRs with mutant template were analyzed. These chromatograms confirm the data shown in Figure 27 and the results taken together clearly demonstrate, that the system generates ligation products in a significant amount only if the wildtype template is provided.

Appropriate control runs were performed to determine retention times of the different compounds involved in the LCR experiments. These include the four oligonucleotides (A, B, C, and D), a synthetic ds 50-mer (with the same sequence as the ligation product), the wildtype template DNA, sonicated salmon sperm DNA and the *Pfu* DNA ligase in ligation buffer.

In order to test which purification procedure should be used before a LCR reaction can be analyzed by MALDI-TOF-MS, aliquots of an unpurified LCR (Figure 30A) and aliquots of the enzyme stock solution (Figure 30B) were analyzed with MALDI-TOF-MS. It turned out that

-94-

appropriate sample preparation is absolutely necessary since all signals in the unpurified LCR correspond to signals obtained in the MALDI-TOF-MS analysis of the *Pfu* DNA ligase. The calculated mass values of oligo A and the ligation product are 7521 Da and 15450 Da, respectively. The data in Figure 30 show that the enzyme solution leads to mass signals which do interfere with the expected signals of the ligation educts and products and therefore makes an unambiguous signal assignment impossible. Furthermore, the spectra showed signals of the detergent Tween20 being part of the enzyme storage buffer which influences the crystallization behavior of the analyte/matrix mixture in an unfavorable way.

In one purification format streptavidin-coated magnetic beads were used. As was shown in a recent paper, the direct desorption of DNA immobilized by Watson-Crick base pairing to a complementary DNA fragment covalently bound to the beads is possible and the non-biotinylated strand will be desorbed exclusively (Tang *et al.* (1995) Nucleic Acids Res. 23:3126-3131). This approach in using immobilized ds DNA ensures that only the non-biotinylated strand will be desorbed. If non-immobilized ds DNA is analyzed both strands are desorbed (Tang *et al.* (1994) Rapid Comm. Mass Spectrom. 7 183-186) leading to broad signals depending on the mass difference of the two single strands. Therefore, employing this system for LCR only the non-ligated oligonucleotide A, with a calculated mass of 7521 Da, and the ligation product from oligo A and oligo B (calculated mass: 15450 Da) will be desorbed if oligo C is biotinylated at the 5'-end and immobilized on streptavidin-coated beads. This results in a simple and unambiguous identification of the LCR educts and products.

Figure 31A shows a MALDI-TOF mass spectrum obtained from two pooled LCRs (performed as described above) purified on streptavidin

-95-

DynaBeads and desorbed directly from the beads showed that the purification method used was efficient (compared with Figure 30). A signal which represents the unligated oligo A and a signal which corresponds to the ligation product could be detected. The agreement
5 between the calculated and the experimentally found mass values is remarkable and allows an unambiguous peak assignment and accurate detection of the ligation product. In contrast, no ligation product but only oligo A could be detected in the spectrum obtained from two pooled LCRs with mutated template (Figure 31B). The specificity and selectivity
10 of the LCR conditions and the sensitivity of the MALDI-TOF detection is further demonstrated when performing the ligation reaction in the absence of a specific template. Figure 32 shows a spectrum obtained from two pooled LCRs in which only salmon sperm DNA was used as a negative control, only oligo A could be detected, as expected.
15 While the results shown in Figure 31A can be correlated to lane 1 of the gel in Figure 27, the spectrum shown in Figure 31B is equivalent to lane 2 in Figure 27, and finally also the spectrum in Figure 32 corresponds to lane 3 in Figure 27. The results are in congruence with the HPLC analysis presented in Figures 28 and 29. While gel
20 electrophoresis (Figure 27) and HPLC (Figures 28 and 29) reveal either an excess or almost equal amounts of ligation product over ligation educts, the analysis by MALDI-TOF mass spectrometry produces a smaller signal for the ligation product (Figure 31A).

The lower intensity of the ligation product signal could be due to
25 different desorption/ionization efficiencies between 24- and a 50-mer. Since the T_m value of a duplex with 50 compared to 24 base pairs is significantly higher, more 24-mer could be desorbed. A reduction in signal intensity can also result from a higher degree of fragmentation in case of the longer oligonucleotides.

-96-

Regardless of the purification with streptavidin DynaBeads, Figure 32 reveals traces of Tween20 in the region around 2000 Da. Substances with a viscous consistence, negatively influence the process of crystallization and therefore can be detrimental to mass spectrometer analysis. Tween20 and also glycerol which are part of enzyme storage buffers therefore should be removed entirely prior to mass spectrometer analysis. For this reason an improved purification procedure which includes an additional ultrafiltration step prior to treatment with DynaBeads was investigated. Indeed, this sample purification resulted in a significant improvement of MALDI-TOF mass spectrometric performance.

Figure 33 shows spectra obtained from two pooled positive (Fig. 33A) and negative (Fig. 33B) LCRs, respectively. The positive reaction was performed with a chemically synthesized, single strand 50mer as template with a sequence equivalent to the ligation product of oligo C and D. Oligo C was 5'-biotinylated. Therefore the template was not detected. As expected, only the ligation product of Oligo A and B (calculated mass 15450 Da) could be desorbed from the immobilized and ligated oligo C and D. This newly generated DNA fragment is represented by the mass signal of 15448 Da in Figure 33A. Compared to Figure 32A, this spectrum clearly shows that this method of sample preparation produces signals with improved resolution and intensity.

EXAMPLE 7

Mutation detection by solid phase oligo base extension of a primer and analysis by MALDI-TOF mass spectrometry (Primer Oligo Base Extension = Probe)

Summary

The solid-phase oligo base extension method detects point mutations and small deletions as well as small insertions in amplified DNA. The method is based on the extension of a detection primer that

-97-

anneals adjacent to a variable nucleotide position on an affinity-captured amplified template, using a DNA polymerase, a mixture of three dNTPs, and the missing one dideoxy nucleotide. The resulting products are evaluated and resolved by MALDI-TOF mass spectrometry without

5 further labeling procedures. The aim of the following experiment was to determine mutant and wildtype alleles in a fast and reliable manner.

Description of the experiment

The method used a single detection primer followed by a oligonucleotide extension step to give products differing in length by

10 some bases specific for mutant or wildtype alleles which can be easily resolved by MALDI-TOF mass spectrometry. The method is described by using as example the exon 10 of the CFTR-gene. Exon 10 of this gene bears the most common mutation in many ethnic groups ($\Delta F508$) that leads in the homozygous state to the clinical phenotype of cystic fibrosis.

15 MATERIALS AND METHODS

Genomic DNA

Genomic DNA were obtained from healthy individuals, individuals homozygous or heterozygous for the $\Delta F508$ mutation, and one individual heterozygous for the 1506S mutation. The wildtype and mutant alleles

20 were confirmed by standard Sanger sequencing.

PCR amplification of exon 10 of the CFTR gene

The primers for PCR amplification were CFEx10-F (5'-GCAAGTGAATCCTGAGCGTG-3' (SEQ ID No. 13) located in intron 9 and biotinylated) and CFEx10-R (5'-GTGTGAAGGGCGTG-3' SEQ ID No. 14)

25 located in intron 10). Primers were used in a concentration of 8 pmol. Taq-polymerase including 10x buffer were purchased from Boehringer-Mannheim and dNTPs were obtained from Pharmacia. The total reaction volume was 50 μ l. Cycling conditions for PCR were initially

-98-

5 min. at 95°C, followed by 1 min. at 94°C, 45 sec at 53°C, and 30 sec at 72°C for 40 cycles with a final extension time of 5 min at 72°C.

Purification of the amplified products

- Amplification products were purified by using Qiagen's PCR purification kit (No. 28106) according to manufacturer's instructions. The elution of the purified products from the column was done in 50 µl TE-buffer (10mM Tris, 1 mM EDTA, pH 7,5).

Affinity-capture and denaturation of the double stranded DNA

- 10 µL aliquots of the purified amplified product were transferred to one well of a streptavidin-coated microtiter plate (No. 1645684 Boehringer-Mannheim or No. 95029262 Labsystems). Subsequently, 10 µl incubation buffer (80 mM sodium phosphate, 400 mM NaCl, 0,4% Tween20, pH 7,5) and 30 µl water were added. After incubation for 1 hour at room temperature the wells were washed three times with 200µl washing buffer (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8). To denature the double stranded DNA the wells were treated with 100 µl of a 50 mM NaOH solution for 3 min and the wells washed three times with 200 µl washing buffer.

Oligo base extension reaction

- 20 The annealing of 25 pmol detection primer (CF508: 5'-CTATATTCATCATAGGAAACACCA-3' (SEQ ID No. 15) was performed in 50 µl annealing buffer (20 mM Tris, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₂, 1% Triton X-100, pH 8) at 50°C for 10 min. The wells were washed three times with 200 µl washing buffer and once in 200 µl TE buffer. The extension reaction was performed by using some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. The total reaction volume was 45 µl, containing of 21 µl water, 6 µl Sequenase-buffer, 3 µl 10 mM DTT solution, 4,5 µl, 0,5 mM of three dNTPs, 4,5 µl, 2 mM the missing one

-99-

ddNTP, 5,5 μ l glycerol enzyme dilution buffer, 0,25 μ l Sequenase 2.0, and 0,25 pyrophosphatase. The reaction was pipetted on ice and then incubated for 15 min at room temperature and for 5 min at 37°C.

Hence, the wells were washed three times with 200 μ l washing buffer

- 5 and once with 60 μ l of a 70 mM NH_4 -Citrate solution.

Denaturation and precipitation of the extended primer

The extended primer was denatured in 50 μ l 10%-DMSO (dimethylsulfoxide) in water at 80°C for 10 min. For precipitation, 10 μ l NH_4 -Acetate (pH 6.5), 0,5 μ l glycogen (10 mg/ml water, Sigma No.

- 10 G1765), and 100 μ l absolute ethanol were added to the supernatant and incubated for 1 hour at room temperature. After centrifugation at 13.000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 1 μ l 18 Mohm/cm H_2O water.

Sample preparation and analysis on MALDI-TOF mass

- 15 *spectrometry*

Sample preparation was performed by mixing 0,3 μ l of each of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$) and of resuspended DNA/glycogen pellet on a sample target and allowed to air dry. Up to 20 samples were

- 20 spotted on a probe target disk for introduction into the source region of an unmodified Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular mass ($M_r(\text{calc})$) were calculated from atomic compositions; reported
25 experimental M_r ($M_r(\text{exp})$) values are those of the singly-protonated form, determined using external calibration.

RESULTS

The aim of the experiment was to develop a fast and reliable method independent of exact stringencies for mutation detection that

-100-

leads to high quality and high throughput in the diagnosis of genetic diseases. Therefore a special kind of DNA sequencing (oligo base extension of one mutation detection primer) was combined with the evaluation of the resulting mini-sequencing products by matrix-assisted
5 laser desorption ionization (MALDI) mass spectrometry (MS). The time-of-flight (TOF) reflectron arrangement was chosen as a possible mass measurement system. To prove this hypothesis, the examination was performed with exon 10 of the CFTR-gene, in which some mutations could lead to the clinical phenotype of cystic fibrosis, the most common
10 monogenetic disease in the Caucasian population.

The schematic presentation as given in Figure 34 shows the expected short sequencing products with the theoretically calculated molecular mass of the wildtype and various mutations of exon 10 of the CFTR-gene (SEQ ID No. 132). The short sequencing products were
15 produced using either ddTTP (Figure 34A; SEQ ID Nos. 133-135) or ddCTP (Figure 34B; SEQ ID Nos. 136-139) to introduce a definitive sequence related stop in the nascent DNA strand. The MALDI-TOF-MS spectra of healthy, mutation heterozygous, and mutation homozygous individuals are presented in Figure 35. All samples were confirmed by
20 standard Sanger sequencing which showed no discrepancy in comparison to the mass spec analysis. The accuracy of the experimental measurements of the various molecular masses was within a range of minus 21.8 and plus 87.1 dalton (Da) to the range expected. This allows a definitive interpretation of the results in each case. A further
25 advantage of this procedure is the unambiguous detection of the Δ I507 mutation. In the ddTTP reaction, the wildtype allele would be detected, whereas in the ddCTP reaction the three base pair deletion would be disclosed.

-101-

- The method described is highly suitable for the detection of single point mutations or microlesions of DNA. Careful choice of the mutation detection primers will open the window of multiplexing and lead to a high throughput including high quality in genetic diagnosis without any need
- 5 for exact stringencies necessary in comparable allele-specific procedures. Because of the uniqueness of the genetic information, the oligo base extension of mutation detection primer is applicable in each disease gene or polymorphic region in the genome like variable number of tandem repeats (VNTR) or other single nucleotide polymorphisms (e.g.,
- 10 apolipoprotein E gene), as also described here.

EXAMPLE 8

Detection of Polymerase Chain Reaction Products Containing 7-Deazapurine Moieties with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry

15 MATERIALS AND METHODS

Nucleic acid amplifications

- The following oligodeoxynucleotide primers were either synthesized according to standard phosphoramidite chemistry (Sinha, N.D., et al., (1983) *Tetrahedron Let.* Vol. 24, Pp. 5843-5846; Sinha,
- 20 N.D., et al., (1984) *Nucleic Acids Res.*, Vol. 12, Pp. 4539-4557) on a MilliGen 7500 DNA synthesizer (Millipore, Bedford, MA, USA) in 200 nmol scales or purchased from MWG-Biotech (Ebersberg, Germany, primer 3) and Biometra (Goettingen, Germany, primers 6-7).
- primer 1: 5'-GTCACCCTCGACCTGCAG (SEQ ID NO. 16);
- 25 primer 2: 5'-TTGTAAAACGACGGCCAGT (SEQ ID NO. 17);
- primer 3: 5'-CTTCCACCGCGATGTTGA (SEQ ID NO. 18);
- primer 4: 5'-CAGGAAACAGCTATGAC (SEQ ID NO. 19);
- primer 5: 5'-GTAAAACGACGGCCAGT (SEQ ID NO. 20);
- primer 6: 5'-GTCACCCTCGACCTGCAGC (g: RiboG) (SEQ ID NO. 21);

-102-

primer 7: 5'-GTTGTAAAACGAGGGCCAgT (g: RiboG) (SEQ ID NO. 22);

The 99-mer (SEQ ID No. 141) and 200-mer DNA strands (SEQ ID No. 140; modified and unmodified) as well as the ribo- and

- 5 7-deaza-modified 100-mer were amplified from pRFc1 DNA (10 ng, generously supplied by S. Feyerabend, University of Hamburg) in 100 μ L reaction volume containing 10 mmol/L KCl, 10 mmol/L $(\text{NH}_4)_2\text{SO}_4$, 20 mmol/L Tris HCl (pH 8.8), 2 mmol/L MgSO_4 , (exo(-) *Pseudococcus furiosus* (Pfu) -Buffer, Pharmacia, Freiburg, Germany), 0.2 mmol/L each
- 10 dNTP (Pharmacia, Freiburg, Germany), 1 μ mol/L of each primer and 1 unit of exo(-)Pfu DNA polymerase (Stratagene, Heidelberg, Germany). For the 99-mer primers 1 and 2, for the 200-mer primers 1 and 3 and for the 100-mer primers 6 and 7 were used. To obtain 7-deazapurine modified nucleic acids, during PCR-amplification dATP and dGTP were
- 15 replaced with 7-deaza-dATP and 7-deaza- dGTP. The reaction was performed in a thermal cycler (OmniGene, MWG-Biotech, Ebersberg, Germany) using the cycle: denaturation at 95°C for 1 min., annealing at 51°C for 1 min. and extension at 72°C for 1 min. For all PCRs the number of reaction cycles was 30. The reaction was allowed to extend
- 20 for additional 10 min. at 72°C after the last cycle.

- The 103-mer DNA strands (modified and unmodified; SEQ ID No. 245) were amplified from M13mp18 RFI DNA (100 ng, Pharmacia, Freiburg, Germany) in 100 μ L reaction volume. using primers 4 and 5 all other concentrations were unchanged. The reaction was performed
- 25 using the cycle: denaturation at 95°C for 1 min., annealing at 40°C for 1 min. and extension at 72°C for 1 min. After 30 cycles for the unmodified and 40 cycles for the modified 103-mer respectively, the samples were incubated for additional 10 min. at 72°C.

Synthesis of 5'-[³²-P]-labeled PCR-primers

-103-

Primers 1 and 4 were 5'-[³²P]-labeled employing T4-polynucleotidkinase (Epicentre Technologies) and (γ -³²P)-ATP (BLU/NGG/502A, Dupont, Germany) according to the protocols of the manufacturer. The reactions were performed substituting 10% of primer 5 1 and 4 in PCR with the labeled primers under otherwise unchanged reaction-conditions. The amplified DNAs were separated by gel electrophoresis on a 10% polyacrylamide gel. The appropriate bands were excised and counted on a Packard TRI-CARB 460C liquid scintillation system (Packard, CT, USA).

10 *Primer-cleavage from ribo-modified PCR-product*

The amplified DNA was purified using Ultrafree-MC filter units (30,000 NMWL), it was then redissolved in 100 μ l of 0.2 mol/L NaOH and heated at 95°C for 25 minutes. The solution was then acidified with HCl (1 mol/L) and further purified for MALDI-TOF analysis employing 15 Ultrafree-MC filter units (10,000 NMWL) as described below.

Purification of amplified products

All samples were purified and concentrated using Ultrafree-MC units 30000 NMWL (Millipore, Eschborn, Germany) according to the manufacturer's description. After lyophilization, amplified products were 20 redissolved in 5 μ L (3 μ L for the 200-mer) of ultrapure water. This analyte solution was directly used for MALDI-TOF measurements.

MALDI-TOF MS

Aliquots of 0.5 μ L of analyte solution and 0.5 μ L of matrix solution (0.7 mol/L 3-HPA and 0.07 mol/L ammonium citrate in acetonitrile/water 25 (1:1, v/v)) were mixed on a flat metallic sample support. After drying at ambient temperature the sample was introduced into the mass spectrometer for analysis. The MALDI-TOF mass spectrometer used was a Finnigan MAT Vision 2000 (Finnigan MAT, Bremen, Germany). Spectra were recorded in the positive ion reflector mode with a 5 keV ion source

-104-

and 20 keV postacceleration. The instrument was equipped with a nitrogen laser (337 nm wavelength). The vacuum of the system was $3\text{-}4\cdot 10^{-8}$ hPa in the analyzer region and $1\text{-}4\cdot 10^{-7}$ hPa in the source region. Spectra of modified and unmodified DNA samples were obtained with the same relative laser power; external calibration was performed with a mixture of synthetic oligodeoxynucleotides (7-to 50-mer).

RESULTS AND DISCUSSION

Enzymatic synthesis of 7-deazapurine nucleotide containing nucleic acids by PCR

10 In order to demonstrate the feasibility of MALDI-TOF MS for the rapid, gel-free analysis of short amplified products and to investigate the effect of 7-deazapurine modification of nucleic acids under MALDI-TOF conditions, two different primer-template systems were used to synthesize DNA fragments. Sequences are displayed in Figures 36 and 15 37. While the two single strands of the 103-mer amplified product had nearly equal masses ($\Delta m = 8$ u), the two single strands of the 99-mer differed by 526 u. Considering that 7-deaza purine nucleotide building blocks for chemical DNA synthesis are approximately 160 times more expensive than regular ones (Product Information, Glen Research Corporation, Sterling, VA) and their application in standard β - 20 cyano-phosphoramidite chemistry is not trivial (Product Information, Glen Research Corporation, Sterling, VA; Schneider *et al.* (1995) Nucl. Acids Res. 23:1570) the cost of 7-deaza purine modified primers would be very high. Therefore, to increase the applicability and scope of the method, 25 all PCRs were performed using unmodified oligonucleotide primers which are routinely available. Substituting dATP and dGTP by c^7 -dATP and c^7 -dGTP in polymerase chain reaction led to products containing approximately 80% 7-deaza-purine modified nucleosides for the 99-mer

-105-

and 103-mer; and about 90% for the 200-mer, respectively. Table II shows the base composition of all PCR products.

TABLE II:

5 Base composition of the 99-mer, 103-mer and 200-mer PCR amplification products (unmodified and 7-deaza purine modified)

	DNA-fragments ¹	C	T	A	G	c ⁷ -deaza-A	c ⁷ -deaza-G	rel. mod. 2
	200-mers	54	34	56	56	-	-	-
	modified 200-mer s	54	34	6	5	50	51	90%
10	200-mer a	56	56	34	54	-	-	-
	modified 200-mer a	56	56	3	4	31	50	92%
	103-mer s	28	23	24	28	-	-	-
	modified 103-mer s	28	23	6	5	18	23	79%
	103-mer a	28	24	23	28	-	-	-
15	modified 103-mer a	28	24	7	4	16	24	78%
	99-mer s	34	21	24	20	-	-	-
	modified 99-mer s	34	21	6	5	18	15	75%
	99-mer a	20	24	21	34	-	-	-
20	modified 99-mer a	20	24	3	4	18	30	87%

¹ "s" and "a" describe "sense" and "antisense" strands of the double-stranded amplified product.

² indicates relative modification as percentage of 7-deaza purine modified nucleotides of total amount of purine nucleotides.

25

It remained to be determined whether 80-90% 7-deaza-purine modification is sufficient for accurate mass spectrometer detection. It was therefore important to determine whether all purine nucleotides could be substituted during the enzymatic amplification step. This was not trivial since it had been shown that c⁷-dATP cannot fully replace dATP in PCR if *Taq* DNA polymerase is employed (Seela, F. and A. Roelling (1992) Nucleic Acids Res., 20,55-61). Fortunately it was found

30

-106-

that *exo(-)Pfu* DNA polymerase indeed could accept c^7 -dATP and c^7 -dGTP in the absence of unmodified purine nucleoside triphosphates. The incorporation was less efficient leading to a lower yield of amplified product (Figure 38).

- 5 To verify these results, the amplifications with [32 P]-labeled primers were repeated. The autoradiogram (Figure 39) clearly shows lower yields for the modified PCR-products. The bands were excised from the gel and counted. For all amplified products the yield of the modified nucleic acids was about 50%, referring to the corresponding unmodified
- 10 amplification product. Further experiments showed that *exo(-)DeepVent* and *Vent* DNA polymerase were able to incorporate c^7 -dATP and c^7 -dGTP during PCR as well. The overall performance, however, turned out to be best for the *exo(-)Pfu* DNA polymerase giving least side products during amplification. Using all three polymerases, it was found that such PCRs
- 15 employing c^7 -dATP and c^7 -dGTP instead of their isosteres showed less side-reactions giving a cleaner PCR-product. Decreased occurrence of amplification side products may be explained by a reduction of primer mismatches due to a long template which is synthesized during PCR. Decreased melting point for DNA duplexes containing 7-deaza-purine
- 20 have been described (Mizusawa, S. et al., (1986) *Nucleic Acids Res.*, 14, 1319-1324). In addition to the three polymerases specified above (*exo(-)Deep Vent* DNA polymerase, *5Vent* DNA polymerase and *exo(-) (Pfu)* DNA polymerase), it is anticipated that other polymerases, such as the Large Klenow fragment of *E.coli* DNA polymerase, *Sequenase*, *Taq* DNA
- 25 polymerase and *U AmpliTaq* DNA polymerase can be used. In addition, where RNA is the template, RNA polymerases, such as the SP6 or the T7 RNA polymerase, must be used.

-107-

MALDI-TOF mass spectrometry of modified and unmodified amplified products.

The 99-mer, 103-mer and 200-mer amplified products were analyzed by MALDI-TOF MS. Based on past experience, it was known that the degree of depurination depends on the laser energy used for desorption and ionization of the analyte. Since the influence of 7-deazapurine modification on fragmentation due to depurination was to be investigated, all spectra were measured at the same relative laser energy.

- 10 Figures 40a and 40b show the mass spectra of the modified and unmodified 103-mer nucleic acids. In case of the modified 103-mer, fragmentation causes a broad $(M+H)^+$ signal. The maximum of the peak is shifted to lower masses so that the assigned mass represents a mean value of $(M+H)^+$ signal and signals of fragmented ions, rather than the
- 15 $(M+H)^+$ signal itself. Although the modified 103-mer still contains about 20% A and G from the oligonucleotide primers, it shows less fragmentation which is featured by much more narrow and symmetric signals. Especially peak tailing on the lower mass side due to depurination, is substantially reduced. Hence, the difference between
- 20 measured and calculated mass is strongly reduced although it is still below the expected mass. For the unmodified sample a $(M+H)^+$ signal of 31670 was observed, which is a 97 u or 0.3% difference to the calculated mass. While, in case of the modified sample this mass difference diminished to 10 u or 0.03% (31713 u found, 31723 u
- 25 calculated). These observations are verified by a significant increase in mass resolution of the $(M+H)^+$ signal of the two signal strands ($n/\Delta m = 67$ as opposed to 18 for the unmodified sample with $\Delta m =$ full width at half maximum, fwhm). Because of the low mass difference between the two single strands (8 u) their individual signals were not resolved.

-108-

With the results of the 99 base pair DNA fragments the effects of increased mass resolution for 7-deazapurine containing DNA becomes even more evident. The two single strands in the unmodified sample were not resolved even though the mass difference between the two strands of the amplified product was very high with 526 u due to unequal distribution of purines and pyrimidines (figure 41a). In contrast to this, the modified DNA showed distinct peaks for the two single strands (figure 41b) which demonstrates the superiority of this approach for the determination of molecular weights to gel electrophoretic methods even more profound. Although base line resolution was not obtained the individual masses were able to be assigned with an accuracy of 0.1%: $\Delta m = 27$ u for the lighter (calc. mass = 30224 u) and $\Delta m = 14$ u for the heavier strand (calc. mass = 30750 u). Again, it was found that the full width at half maximum was substantially decreased for the 7-deazapurine containing sample.

In case the 99-mer and 103-mer, the 7-deazapurine containing nucleic acids seem to give higher sensitivity despite the fact that they still contain about 20% unmodified purine nucleotides. To get comparable signal-to-noise ratio at similar intensities for the $(M+H)^+$ signals, the unmodified 99-mer required 20 laser shots in contrast to 12 for the modified one and the 103-mer required 12 shots for the unmodified sample as opposed to three for the 7-deazapurine nucleoside-containing amplified product.

Comparing the spectra of the modified and unmodified 200-mer amplicons, improved mass resolution was again found for the 7-deazapurine containing sample as well as increased signal intensities (Figures 42A and 42B). While the signal of the single strands predominates in the spectrum of the modified sample the DNA-duplex

-109-

and dimers of the single strands gave the strongest signal for the unmodified sample.

A complete 7-deaza purine modification of nucleic acids may be achieved either using modified primers in PCR or cleaving the unmodified
5 primers from the partially modified amplified product. Since disadvantages are associated with modified primers, as described above, a 100-mer was synthesized using primers with a ribo-modification. The primers were cleaved hydrolytically with NaOH according to a method developed earlier in our laboratory (Koester, H. et al., *Z Physiol. Chem.*,
10 359, 1570-1589). Figures 43A and 43B display the spectra of the amplified product before and after primer cleavage. Figure 43b shows that the hydrolysis was successful: The hydrolyzed amplified product as well as the two released primers could be detected together with a small signal from residual uncleaved 100-mer. This procedure is especially
15 useful for the MALDI-TOF analysis of very short PCR-products since the share of unmodified purines originating from the primer increases with decreasing length of the amplified sequence.

The remarkable properties of 7-deazapurine modified nucleic acids can be explained by either more effective desorption and/or ionization,
20 increased ion stability and/or a lower denaturation energy of the double stranded purine modified nucleic acid. The exchange of the N-7 for a methyl group results in the loss of one acceptor for a hydrogen bond which influences the ability of the nucleic acid to form secondary structures due to non-Watson-Crick base pairing (Seela, F. and A. Kehne
25 (1987) *Biochemistry*, 26, 2232-2238.). In addition to this the aromatic system of 7-deazapurine has a lower electron density that weakens Watson-Crick base pairing resulting in a decreased melting point (Mizusawa, S. et al., (1986) *Nucleic Acids Res.*, 14, 1319-1324) of the double-strand. This effect may decrease the energy needed for

-110-

denaturation of the duplex in the MALDI process. These aspects as well as the loss of a site which probably will carry a positive charge on the N-7 nitrogen renders the 7-deazapurine modified nucleic acid less polar and may promote the effectiveness of desorption.

- 5 Because of the absence of N-7 as proton acceptor and the decreased polarization of the C-N bond in 7-deazapurine nucleosides depurination following the mechanisms established for hydrolysis in solution is prevented. Although a direct correlation of reactions in solution and in the gas phase is problematic, less fragmentation due to
- 10 depurination of the modified nucleic acids can be expected in the MALDI process. Depurination may either be accompanied by loss of charge which decreases the total yield of charged species or it may produce charged fragmentation products which decreases the intensity of the non fragmented molecular ion signal.
- 15 The observation of increased sensitivity and decreased peak tailing of the $(M+H)^+$ signals on the lower mass side due to decreased fragmentation of the 7-deazapurine containing samples indicate that the N-7 atom indeed is essential for the mechanism of depurination in the MALDI-TOF process. In conclusion, 7-deazapurine containing nucleic
- 20 acids show distinctly increased ion-stability and sensitivity under MALDI-TOF conditions and therefore provide for higher mass accuracy and mass resolution.

EXAMPLE 9

Solid Phase Sequencing and Mass Spectrometer Detection

25 MATERIALS AND METHODS

Oligonucleotides were purchased from Operon Technologies (Alameda, CA) in an unpurified form. Sequencing reactions were performed on a solid surface using reagents from the sequencing kit for Sequenase Version 2.0 (Amersham, Arlington Heights, Illinois).

-111-

Sequencing a 39-mer target

Sequencing complex:

	SEQUENCE	SEQ ID NO.
5	5'-TCTGGCCTGGTGCAGGGCCTATTGTAGTTGTGACGTACA-(A ^b) ₅ -3'	23
	5'-TGTACGTCACAACT-3' (PNA 16/DNA)	24

In order to perform solid-phase DNA sequencing, template strand DNA11683 was 3'-biotinylated by terminal deoxynucleotidyl transferase.

- 10 A 30 μ l reaction, containing 60 pmol of DNA11683, 1.3 nmol of biotin 14-dATP (GIBCO BRL, Grand Island, NY), 30 units of terminal transferase (Amersham, Arlington Heights, Illinois), and 1x reaction buffer (supplied with enzyme), was incubated at 37°C for 1 hour. The reaction was stopped by heat inactivation of the terminal transferase at 70°C for 10
- 15 min. The resulting product was desalted by passing through a TE-10 spin column (Clontech). More than one molecules of biotin-14-dATP could be added to the 3'-end of DNA11683. The biotinylated DNA11683 was incubated with 0.3 mg of Dynal streptavidin beads in 30 μ l 1x binding and washing buffer at ambient temperature for 30 min.
- 20 The beads were washed twice with TE and redissolved in 30 μ l TE, 10 μ l aliquot (containing 0.1 mg of beads) was used for sequencing reactions.

- The 0.1 mg beads from previous step were resuspended in a 10 μ l volume containing 2 μ l of 5x Sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM MgCl₂, and 250 mM NaCl) from the Sequenase kit and 5
- 25 pmol of corresponding primer PNA 16/DNA. The annealing mixture was heated to 70°C and allowed to cool slowly to room temperature over a 20-30 min time period. Then 1 μ l 0.1 M dithiothreitol solution, 1 μ l Mn buffer (0.15 M sodium isocitrate and 0.1 M MgCl₂), and 2 μ l of diluted Sequenase (3.25 units) were added. The reaction mixture was divided

-112-

- into four aliquots of 3 μ l each and mixed with termination mixes (each contains of 3 μ l of the appropriate termination mix: 32 μ M c7dATP, 32 μ M dCTP, 32 μ M c7dGTP, 32 μ M dTTP and 3.2 μ M of one of the four ddTNPs, in 50 mM NaCl). The reaction mixtures were incubated at 37°C
- 5 for 2 min. After the completion of extension, the beads were precipitated and the supernatant was removed. The beads were washed twice and resuspended in TE and kept at 4°C.

Sequencing a 78-mer target

Sequencing complex:

- 10 5'-AAGATCTGACCAGGGATTTCGGTTAGCGTGACTGCTGCTGCTGCTGCTGCTGCTGCTGGATGATCCGACGCATCAGATCTGG-(A^b)_n-3' (SEQ ID NO. 25)
(TNR.PLASM2)
5'-CTGATGCGTCGGATCATC-3' (CM1) (SEQ ID NO. 26)

- The target TNR.PLASM2 was biotinylated and sequenced using
- 15 procedures similar to those described in previous section (sequencing a 39-mer target).

Sequencing a 15-mer target with partially duplex probe

Sequencing complex:

- 5'-F-GATGATCCGACGCATCACAGCTC^{3'} (SEQ ID No. 27)
- 20 5'-TCGGTTCCAAGAGCTGTGATGCGTCGGATCATC-b-3' (SEQ ID No. 28)

- CM1B3B was immobilized on Dynabeads M280 with streptavidin (Dyna, Norway) by incubating 60 pmol of CM1B3B with 0.3 magnetic beads in 30 μ l 1 M NaCl and TE (1x binding and washing buffer) at room temperature for 30 min. The beads were washed twice with TE and
- 25 redissolved in 30 μ l TE, 10 or 20 μ l aliquot (containing 0.1 or 0.2 mg of beads respectively) was used for sequencing reactions.

-113-

- The duplex was formed by annealing corresponding aliquot of beads from previous step with 10 pmol of DF11a5F (or 20 pmol of DF11a5F for 0.2 mg of beads) in a 9 μ l volume containing 2 μ l of 5x Sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, and 250 mM NaCl) from the Sequenase kit. The annealing mixture was heated to 65°C and allowed to cool slowly to 37°C over a 20-30 min time period. The duplex primer was then mixed with 10 pmol of TS10 (20 pmol of TS10 for 0.2 mg of beads) in 1 μ l volume, and the resulting mixture was further incubated at 37°C for 5 min, room temperature for 5-10 min.
- 10 Then 1 μ l 0.1 M dithiothreitol solution, 1 μ l Mn buffer (0.15 M sodium isocitrate and 0.1 M $MnCl_2$), and 2 μ l of diluted Sequenase (3.25 units) were added. The reaction mixture was divided into four aliquots of 3 μ l each and mixed with termination mixes (each contains of 4 μ l of the appropriate termination mix: 16 μ M dATP, 16 μ M dCTP, 16 μ M dGTP, 16 μ M dTTP and 1.6 μ M of one of the four ddNTPs, in 50 mM NaCl). The reaction mixtures were incubated at room temperature for 5 min, and 37°C for 5 min. After the completion of extension, the beads were precipitated and the supernatant was removed. The beads were resuspended in 20 μ l TE and kept at 4°C. An aliquot of 2 μ l (out of 20 μ l) from each tube was taken and mixed with 8 μ l of formamide, the resulting samples were denatured at 90-95°C for 5 min and 2 μ l (out of 10 μ l total) was applied to an ALF DNA sequencer (Pharmacia, Piscataway, NJ) using a 10% polyacrylamide gel containing 7 M urea and 0.6x TBE. The remaining aliquot was used for MALDI-TOF MS analysis.
- 25

-114-

MALDI sample preparation and instrumentation

Before MALDI analysis, the sequencing ladder loaded magnetic beads were washed twice using 50 mM ammonium citrate and resuspended in 0.5 μ l pure water. The suspension was then loaded onto the sample target of the mass spectrometer and 0.5 μ l of saturated matrix solution (3-hydroxypicolinic acid (HPA): ammonium citrate = 10:1 mole ratio in 50% acetonitrile) was added. The mixture was allowed to dry prior to mass spectrometer analysis.

The reflectron TOFMS mass spectrometer (Vision 2000, Finnigan MAT, Bremen, Germany) was used for analysis. 5 kV was applied in the ion source and 20 kV was applied for postacceleration. All spectra were taken in the positive ion mode and a nitrogen laser was used. Normally, each spectrum was averaged for more than 100 shots and a standard 25-point smoothing was applied.

15 RESULTS AND DISCUSSION*Conventional solid-phase sequencing*

In conventional sequencing methods, a primer is directly annealed to the template and then extended and terminated in a Sanger dideoxy sequencing. Normally, a biotinylated primer is used and the sequencing ladders are captured by streptavidin-coated magnetic beads. After washing, the products are eluted from the beads using EDTA and formamide. Previous findings indicated that only the annealed strand of a duplex is desorbed and the immobilized strand remains on the beads. Therefore, it is advantageous to immobilize the template and anneal the primer. After the sequencing reaction and washing, the beads with the immobilized template and annealed sequencing ladder can be loaded directly onto the mass spectrometer target and mix with matrix. In MALDI, only the annealed sequencing ladder will be desorbed and ionized, and the immobilized template will remain on the target.

-115-

A 39-mer template (SEQ ID No. 23) was first biotinylated at the 3'-end by adding biotin-14-dATP with terminal transferase. More than one biotin-14-dATP molecule could be added by the enzyme. Since the template was immobilized and remained on the beads during MALDI, the

5 number of biotin-14-dATP would not affect the mass spectra. A 14-mer primer (SEQ ID No. 24) was used for the solid-state sequencing to generate DNA fragments 3-27 below (SEQ ID Nos. 142-166). MALDI-TOF mass spectra of the four sequencing ladders are shown in Figure 44 and the expected theoretical values are shown in Table III.

-116-

TABLE III

5	1	5'-TCTGGCCTGGTGCAGGGCCTATTGTAGTTGTGACGTACA-(A ⁸) _n -3'
	2	3'-TCAACACTGCATGT-5'
	3	3'-ATCAACACTGCATGT-5'
	4	3'-CATCAACACTGCATGT-5'
	5	3'-ACATCAACACTGCATGT-5'
10	6	3'-AACATCAACACTGCATGT-5'
	7	3'-TAACATCAACACTGCATGT-5'
	8	3'-ATAACATCAACACTGCATGT-5'
	9	3'-GATAACATCAACACTGCATGT-5'
	10	3'-GGATAACATCAACACTGCATGT-5'
15	11	3'-CGGATAACATCAACACTGCATGT-5'
	12	3'-CCGGATAACATCAACACTGCATGT-5'
	13	3'-CCCGGATAACATCAACACTGCATGT-5'
	14	3'-TCCCGGATAACATCAACACTGCATGT-5'
	15	3'-GTCCCGGATAACATCAACACTGCATGT-5'
20	16	3'-CGTCCCGGATAACATCAACACTGCATGT-5'
	17	3'-ACGTCCCGGATAACATCAACACTGCATGT-5'
	18	3'-CACGTCCCGGATAACATCAACACTGCATGT-5'
	19	3'-CCACGTCCCGGATAACATCAACACTGCATGT-5'
	20	3'-ACCACGTCCCGGATAACATCAACACTGCATGT-5'
25	21	3'-GACCACGTCCCGGATAACATCAACACTGCATGT-5'
	22	3'-GGACCACGTCCCGGATAACATCAACACTGCATGT-5'
	23	3'-CGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
	24	3'-CCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
	25	3'-ACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
	26	3'-GACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
	27	3'-AGACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'

-117-

TABLE III (Continued)

	A-reaction	C-reaction	G-reaction	T-reaction
1.				
2.	4223.8	4223.8	4223.8	4223.8
5	3.	4521.1		
4.		4809.2		
5.	5133.4			
6.	5434.6			
7.				5737.8
10	8.	6051.1		
9.			6379.2	
10.			6704.4	
11.		6995.6		
12.		7284.8		
15	13.	7574.0		
14.				7878.2
15.			8207.4	
16.		8495.6		
17.	8808.8			
20	18.	9097.0		
19.		9386.2		
20.	9699.4			
21.			10027.6	
22.			10355.8	
25	23.	10644.0		
24.		10933.2		
25.	11246.4			

-118-

	A-reaction	C-reaction	G-reaction	T-reaction
26.			11574.6	
27.	11886.8			

The sequencing reaction produced a relatively homogenous ladder, and the full-length sequence was determined easily. One peak around 5150 appeared in all reactions are not identified. A possible explanation is that a small portion of the template formed some kind of secondary structure, such as a loop, which hindered sequenase extension. Mis-incorporation is of minor importance, since the intensity of these peaks were much lower than that of the sequencing ladders. Although 7-deaza purines were used in the sequencing reaction, which could stabilize the N-glycosidic bond and prevent depurination, minor base losses were still observed since the primer was not substituted by 7-deazapurines. The full length ladder, with a ddA at the 3' end, appeared in the A reaction with an apparent mass of 11899.8. A more intense peak of 12333 appeared in all four reactions and is likely due to an addition of an extra nucleotide by the Sequenase enzyme.

The same technique could be used to sequence longer DNA fragments. A 78-mer template containing a CTG repeat (SEQ ID No. 25) was 3'-biotinylated by adding biotin-14-dATP with terminal transferase. An 18-mer primer (SEQ ID No. 26) was annealed right outside the CTG repeat so that the repeat could be sequenced immediately after primer extension. The four reactions were washed and analyzed by MALDI-TOFMS as usual. An example of the G-reaction is shown in Figure 45 (SEQ ID Nos. 167-220) and the expected sequencing ladder is shown in Table IV with theoretical mass values for each ladder component. All sequencing peaks were well resolved except the last component (theoretical value 20577.4) was indistinguishable from the

-119-

background. Two neighboring sequencing peaks (a 62-mer and a 63-mer) were also separated indicating that such sequencing analysis could be applicable to longer templates. Again, an addition of an extra nucleotide by the Sequenase enzyme was observed in this spectrum.

- 5 This addition is not template specific and appeared in all four reactions which makes it easy to be identified. Compared to the primer peak, the sequencing peaks were at much lower intensity in the long template case.

TABLE IV

AAGATCTGACCAGGGAATCGGTTAGCGTGACTGCTGCTGATGATCCGACGCATCAGATCTGG-(A ⁸) _n -3'	
1	3'-CTACTAGGCTGCGTAGTC-5'
2	3'-CCTACTAGGCTGCGTAGTC-5'
3	3'-ACCTACTAGGCTGCGTAGTC-5'
4	3'-GACCTACTAGGCTGCGTAGTC-5'
5	3'-CGACCTACTAGGCTGCGTAGTC-5'
6	3'-ACGACCTACTAGGCTGCGTAGTC-5'
7	3'-GACGACCTACTAGGCTGCGTAGTC-5'
8	3'-CGACGACCTACTAGGCTGCGTAGTC-5'
9	3'-ACGACGACCTACTAGGCTGCGTAGTC-5'
10	3'-GACGACGACCTACTAGGCTGCGTAGTC-5'
11	3'-CGACGACGACCTACTAGGCTGCGTAGTC-5'
12	3'-ACGACGACGACCTACTAGGCTGCGTAGTC-5'
13	3'-GACGACGACGACCTACTAGGCTGCGTAGTC-5'
14	3'-CGACGACGACGACCTACTAGGCTGCGTAGTC-5'

15	3'-ACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
16	3'-GACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
17	3'-CGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
18	3'-ACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
19	3'-GACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
20	3'-CGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
21	3'-ACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
22	3'-GACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
23	3'-CGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
24	3'-ACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
25	3'-GACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
26	3'-TGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
27	3'-CTGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
28	3'-ACTGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
29	3'-CACTGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
30	3'-GCACTGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'

5

10

15

[illegible]

53	3'-CTAGACTGGTCCCTAAGCCAAATCGCACTGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
54	3'-TCTAGACTGGTCCCTAAGCCAAATCGCACTGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
55	3'-TTCTAGACTGGTCCCTAAGCCAAATCGCACTGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'

-124-

TABLE IV Continued

	ddATP	ddCTP	ddGTP	ddTTP
5	1. 5491.6	5491.6	5491.6	5491.6
	2.	5764.8		
	3. 6078.0			
	4.		6407.2	
	5.	6696.4		
10	6. 7009.6			
	7.		7338.8	
	8.	7628.0		
	9. 7941.2			
	10.		8270.4	
15	11.	8559.6		
	12. 8872.8			
	13.		9202.0	
	14.	9491.2		
	15. 9804.4			
20	16.		10133.6	
	17.	10422.88		
	18. 10736.0			
	19.		11065.2	
	20.	11354.4		
25	21. 11667.6			
	22.		11996.8	
	23.	12286.0		
	24. 12599.2			

-125-

5	25.		12928.4	
	26.			13232.6
	27.	13521.8		
	28.	13835.0		
	29.	14124.2		
10	30.		14453.4	
	31.	14742.6		
	32.			15046.8
	33.	15360.0		
	34.	15673.2		
15	35.	15962.4		
	36.	16251.6		
	37.		16580.8	
	38.	16894.0		
	39.	17207.2		
20	40.			17511.4
	41.	17800.6		
	42.	18189.8		
	43.	18379.0		
	44.			18683.2
25	45.		19012.4	
	46.		19341.6	
	47.			19645.8
	48.	19935.0		
	49.	20248.2		
	50.		20577.4	

-126-

51.	20890.6			
52.				21194.4
53.		21484.0		
54.				21788.2
55.				22092.4

Sequencing using duplex DNA probes for capturing and priming

Duplex DNA probes with single-stranded overhang have been demonstrated to be able to capture specific DNA templates and also serve as primers for solid-state sequencing. The scheme is shown in Figure 46. Stacking interactions between a duplex probe and a single-stranded template allow only a 5-base overhang to be sufficient for capturing. Based on this format, a 5' fluorescent-labeled 23-mer (5'-GAT GAT CCG ACG CAT CAC AGC TC-3') (SEQ ID No. 29) was annealed to a 3'-biotinylated 18-mer (5'-GTG ATG CGT CGG ATC ATC-3') (SEQ ID No. 30), leaving a 5-base overhang. A 15-mer template (5'-TCG GTT CCA AGA GCT-3') (SEQ ID No. 31) was captured by the duplex and sequencing reactions were performed by extension of the 5-base overhang. MALDI-TOF mass spectra of the reactions are shown in Figure 47A-D. All sequencing peaks were resolved although at relatively low intensities. The last peak in each reaction is due to unspecific addition of one nucleotide to the full length extension product by the Sequenase enzyme. For comparison, the same products were run on a conventional DNA sequencer and a stacking fluorogram of the results is shown in Figure 48. As can be seen from the Figure, the mass spectra had the same pattern as the fluorogram with sequencing peaks at much lower intensity compared to the 23-mer primer.

-127-

EXAMPLE 10

Thermo Sequenase Cycle Sequencing

MATERIALS AND METHODS

- PCR amplification.* Human leukocytic genomic DNA was used for
- 5 PCR amplification. PCR primers to amplify a 209 bp fragment of the β -globin gene were the β 2 forward primer (5'-CAT TTG CTT CTG ACA CAA CTG-3' SEQ ID NO. 32) and the β 11 reverse primer (5'-CTT CTC TGT CTC CAC ATG C-3' SEQ ID NO. 33). Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs
- 10 from Pharmacia (Freiburg, Germany). The total reaction volume was 50 μ l including 8 pmol of each primer with approximately 200 ng of genomic DNA used as template and a final dNTP concentration of 200 μ M. PCR conditions were: 5 min at 94°C, followed by 40 cycles of 30 sec at 94°C, 45 sec at 53°C, 30 sec at 72°C, and a final extension time of 2
- 15 min at 72°C. The generated amplified product was purified and concentrated (2x) with the Qiagen 'Qiaquick' PCR purification kit (#28106) and stored in H₂O.

- Cycle Sequencing. Sequencing ladders were generated by primer extension with Thermo SequenaseTM-DNA Polymerase (Amersham
- 20 LIFE Science, #E79000Y) under the following conditions: 7 pmol of HPLC purified primer (Cod5 12mer: 5'-TGC ACC TGA CTC-3' SEQ ID No. 34) were added to 6 μ l purified and concentrated amplified product (i.e. 12 μ l of the original amplified product), 2.5 units Thermo Sequenase and 2.5 ml Thermo Sequenase reaction buffer in a total volume of 25 μ l. The final
- 25 nucleotide concentrations were 30 μ M of the appropriate ddNTP (ddATP, ddCTP, ddGTP or ddTTP; Pharmacia Biotech, #27-2045-01) and 210 μ M of each dNTP (7-deaza-dATP, DCTP, 7-deaza-GTP, dTTP; Pharmacia Biotech).

-128-

Cycling conditions were: denaturation for 4 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 38°C, 30 sec at 55°C, and a final extension of 2 min at 72°C.

- Sample preparation and analysis by MALDI-TOF MS. After
- 5 completion of the cycling program, the reaction volume was increased to 50 μ l by addition of 25 μ l H₂O. Desalting was achieved by shaking 30 μ l of ammonium saturated DOWEX (Fluka #44485) cation exchange beads with 50 μ l of the analyte for 2 min at room temperature. The Dowex
- 10 NH₄OH to convert them to the ammonium form, then washed with H₂O until the supernatant was neutral, and finally put in 10 mM ammonium citrate for usage. After the cation exchange, DNA was purified and concentrated by ethanol precipitation by adding 5 μ l 3 M ammonium
- 15 acetate (pH 6.5), 0.5 μ l glycogen (10 mg/ml, Sigma), and 110 μ l absolute ethanol to the analyte and incubated at room temperature for 1 hour. After 12 min centrifugation at 20,000 X g the pellet was washed in 70% ethanol and resuspended in 1 μ l 18 Mohm/cm H₂O water.

- For MALDI-TOF MS analysis 0.35 μ l of resuspended DNA was mixed with 0.35-1.3 μ l matrix solution (0.7 M 3-hydroxypicolinic acid
- 20 (3-HPA), 0.07 M ammonium citrate in 1:1 H₂O:CH₃CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion
- 25 dynode, respectively. External calibration generated from eight peaks (3000-18000 Da) was used for all spectra.

RESULTS

FIGURE 49 shows a MALDI-TOF mass spectrum of the sequencing ladder generated from a biological amplified product as template and a 12mer (5'-TGC ACC TGA CTC-3' (SEQ ID NO.34)) sequencing primer.

-129-

The peaks resulting from depurinations and peaks which are not related to the sequence are marked by an asterisk. MALDI-TOF MS measurements were taken on a reflectron TOF MS. A.) Sequencing ladder stopped with ddATP; B.) Sequencing ladder stopped with ddCTP; 5 C.) Sequencing ladder stopped with ddGTP; D.) Sequencing ladder stopped with ddTTP.

FIGURE 50 shows a schematic representation of the sequencing ladder generated in Fig. 49 with the corresponding calculated molecular masses up to 40 bases after the primer (SEQ ID Nos 221-260). For the 10 calculation the following masses were used: 3581.4 Da for the primer, 312.2 Da for 7-deaza-dATP, 304.2 Da for dTTP, 289.2 Da for dCTP and 328.2 Da for 7-deaza-dGTP.

FIGURE 51 shows the sequence of the amplified 209bp amplified product within the β -globin gene (SEQ ID No. 261), which was used as a 15 template for sequencing. The sequences of the appropriate PCR primer and the location of the 12mer sequencing primer is also shown. This sequence represents a homozygote mutant at the position 4 after the primer. In a wildtype sequence this T would be replaced by an A.

EXAMPLE 11

20 Microsatellite Analysis Using Primer Oligo Base Extension (PROBE) and MALDI-TOF Mass Spectrometry

SUMMARY

The method uses a single detection primer followed by an oligonucleotide extension step to give products differing in length by a 25 number of bases specific for the number of repeat units or for second site mutations within the repeated region, which can be easily resolved by MALDI-TOF mass spectrometry. The method is demonstrated using as a model system the AluVpA polymorphism in intron 5 of the interferon- α receptor gene located on human chromosome 21, and the

-130-

poly T tract of the splice acceptor site of intron 8 from the CFTR gene located on human chromosome 7.

MATERIALS AND METHODS

Genomic DNA was obtained from 18 unrelated individuals and one family including of a mother, father, and three children. The repeated region was evaluated conventionally by denaturing gel electrophoresis and results obtained were confirmed by standard Sanger sequencing.

The primers for PCR amplification (8 pmol each) were IFNAR-IVS5-5': (5'-TGC TTA CTT AAC CCA GTG TG-3' SEQ ID. NO.35) and IFNAR-IVS5-3'.2: (5'-CAC ACT ATG TAA TAC TAT GC-3' SEQ ID. NO.36) for a part of the intron 5 of the interferon- α receptor gene, and CFEx9-F:(5'-GAA AAT ATC TGA CAA ACT CAT C-3' SEQ ID. NO.37) (5'-biotinylated) and CFEx9-R:(5'-CAT GGA CAC CAA ATT AAG TTC-3' SEQ ID. NO.38) for CFTR exon 9 with flanking intron sequences of the CFTR gene. Taq-polymerase including 10x buffer were purchased from Boehringer-Mannheim and dNTPs were obtained from Pharmacia. The total reaction volume was 50 μ l. PCR conditions were 5 min at 94°C followed by 40 cycles of: 1 min at 94°C, 45 sec at 53°C, and 30 sec at 72°C, and a final extension time of 5 min at 72°C.

Amplification products were purified using Qiagen's PCR purification kit (No.28106) according to manufacturer's instructions. Purified products were eluted from the column in 50 μ l TE-buffer (10mM Tris-HCl, 1 mM EDTA, pH 7.5).

A) Primer oligo base extension reaction (thermo cycling method) CyclePROBE was performed with 5 pmol appropriate detection primer (IFN:5'-TGA GAC TCT GTC TC-3' SEQ ID. NO.39) in a total volume of 25 μ l including 1 pmol purified template, 2 units Thermosequenase (Amersham Life Science, Cat. #E79000Y) 2.5 μ l Thermosequenase buffer, 25 μ mol of each deoxynucleotide

-131-

(7-deaza-dATP, dTTP, and in some experiments extra dCTP) and 100 μ mol of dideoxyguanine and in some experiments additional ddCTP. Cycling conditions: initial denaturation 94°C for 5 min followed by 30 cycles with 44°C annealing temperature for 30 sec and 55°C extension temperature for 1 min.

Primer oligo base extension reaction (isothermal method)

- 10 μ l aliquots of the purified double-stranded amplified product (~3 pmol) were transferred to a streptavidin-coated microliter plate well (~16 pmol capacity per 50 μ l volume; No. 1645684 (Boehringer-Mannheim), followed by addition of 10 μ l incubation buffer (80 mM sodium phosphate, 400 mM NaCl, 0.4% Tween 20, pH 7.5) and 30 μ l water. After incubation for 1 hour at room temperature, the wells were washed three times with 200 μ l washing buffer A (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8) and incubated with 15 100 μ l of 50 mM NaOH for 3 min to denature the double-stranded DNA. Finally, the wells were washed three times with 200 μ l 70 mM ammonium citrate solution.

- The annealing of 100 pmol detection primer (CFpT: 5'-TTC CCC AAA TCC CTG-3' SEQ ID NO. 40) was performed in 50 μ l annealing buffer (50 mM ammonium phosphate buffer, pH 7.0 and 100 mM ammonium chloride) at 65°C for 2 min, at 37°C for 10 min, and at room temperature for 10 min. The wells were washed three times with 200 μ l washing buffer B (40 mM Tris, 1 mM EDTA, 50 mM NH₄Cl, 0.1% Tween 20, pH 8.8) and once in 200 μ l TE buffer. The extension reaction was 25 performed using some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. Total reaction volume was 45 μ l, containing of 21 μ l water, 6 μ l Sequenase-buffer, 3 μ l 100 mM DTT solution, 50 μ mol of 7-deaza-dATP, 20 μ mol ddCTP, 5.5 μ l glycerol enzyme dilution buffer, 0.25 μ l Sequenase 2.0, and 0.25 μ l

-132-

pyrophosphatase. The reaction was pipetted on ice and incubated for 15 min at room temperature and for 5 min at 37°C. Finally, the wells were washed three times with 200 μ l washing buffer B.

- The extended primer was denatured from the template strand by
- 5 heating at 80°C for 10 min in 50 μ l of a 50 mM ammonium hydroxide solution.

For precipitation, 10 μ l 3 M NH_4 -acetate (pH 6.5), 0.5 μ l glycogen (10 mg/ml water, Sigma, Cat.#G1765), and 110 μ l absolute ethanol were added to the supernatant and incubated for 1 hour at room temperature.

- 10 After centrifugation at 13.000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 1 μ l 18 Mohm/cm H_2O water.

- Sample preparation was performed by mixing 0.6 μ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$) with 0.3 μ l of resuspended DNA/glycogen
- 15 pellet on a sample target and allowed to air dry. Up to 20 samples were spotted on a probe target disk for introduction into the source region of a Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular mass
- 20 ($M_r(\text{calc})$) were calculated from atomic compositions; reported experimental M_r ($M_r(\text{exp})$) values are those of the singly-protonated form, determined using external calibration.

RESULTS

- The aim of the experiments was to develop a fast and reliable
- 25 method for the exact determination of the number of repeat units in microsatellites or the length of a mononucleotide stretch including the potential to detect second site mutations within the polymorphic region. Therefore, a special kind of DNA sequencing (primer oligo base extension, PROBE) was combined with the evaluation of the resulting

-133-

products by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS). The time-of-flight (TOF) reflectron arrangement was chosen-as a possible mass measurement system. As an initial feasibility study, an examination was performed first on an AluVpA repeat

- 5 polymorphism located in intron 5 of the human interferon- α receptor gene (cyclePROBE reaction) and second on the poly T tract located in intron 8 of the human CFTR gene (isothermal PROBE reaction).

- A schematic presentation of the cyclePROBE experiment for the AluVpA repeat polymorphism is given in Figure 52. The extension of the
10 antisense strand (SEQ ID No. 262) was performed with the sense strand serving as the template. The detection primer is underlined. In a family study co-dominant segregation of the various alleles could be demonstrated by the electrophoretic procedure as well as by the cyclePROBE method followed by mass spec analysis (Figure 53). Those
15 alleles of the mother and child 2, for which direct electrophoresis of the amplified product indicated one of the two copies to have 13 repeat units, were measured using cyclePROBE to have instead only 11 units using ddG as terminator. The replacement of ddG by ddC resulted in a further unexpected short allele with a molecular mass of approximately
20 11650 in the DNA of the mother and child 2 (Figure 54). Sequence analysis verified this presence of two second site mutations in the allele with 13 repeat units. The first is a C to T transition in the third repeat unit and the second mutation is a T to G transversion in the ninth repeat unit. Examination of 28 unrelated individuals shows that the 13 unit
25 allele is spliced into a normal allele and a truncated allele using cyclePROBE. Statistical evaluation shows that the polymorphism is in Hardy-Weinberg equilibrium for both methods, however, using cyclePROBE as detection method the polymorphism information content is increased to 0.734.

-134-

PROBE was also used as an isothermic method for the detection of the three common alleles at the intron 8 splice acceptor site of the CFTR gene (SEQ ID No. 263). Figure 55 shows a schematic presentation of the expected diagnostic products (SEQ ID Nos. 264-266) with the

5 theoretical mass values. The reaction was also performed in the antisense direction.

Figure 56 demonstrates that all three common alleles (T5, T7, and T9, respectively) at this locus could be reliably disclosed by this method. Reference to Figure 56 indicates that mass accuracy and precision with

10 the reflectron time of flight used in this study ranged from 0-0.4%, with a relative standard deviation of 0.13%. This corresponds to far better than single base accuracy for the up to <90-mer diagnostic products generated in the IFNAR system. Such high analytical sensitivity is sufficient to detect single or multiple insertion/deletion mutations within

15 the repeat unit or its flanking regions, which would induce >1% mass shifts in a 90-mer. This is analogous to the Figure 56 polyT tract analysis. Other mutations (i.e. an A to T or a T to A mutation within the IFNAR gene A3T repeat) which do not cause premature product termination are not detectable using any dNTP/ddNTP combination with

20 PROBE and low performance MS instrumentation; a 9 Da shift in a 90-mer corresponds to a 0.03% mass shift. Achieving the accuracy and precision required to detect such minor mass shifts has been demonstrated with higher performance instrumentation such as Fourier transform (FT)MS, for which single Da accuracy is obtained up to

25 100-mers. Further, tandem FTMS, in which a mass shifted fragment can be isolated within the instrument and dissociated to generate sequence specific fragments, has been demonstrated to locate point mutations to the base in comparably sized products. Thus the combination of PROBE with higher performance instrumentation will have an analytical

-135-

sensitivity which can be matched only by cumbersome full sequencing of the repeat region.

EXAMPLE 12

5 Improved Apolipoprotein E Genotyping Using Primer Oligo Base Extension (PROBE) and MALDI-TOF Mass Spectrometry

MATERIALS AND METHODS

PCR amplification.

Human leukocytic genomic DNA from 100 anonymous individuals from a previously published study (Braun, A *et al.*, (1992) Human Genet. 10 89:401-406) were screened for apolipoprotein E genotypes using conventional methods. PCR primers to amplify a portion of exon 4 of the apo E gene were delineated according to the published sequence (Das, HK *et al.*, (1985) J. Biol. Chem. 260:6240-6247) (forward primer, apoE-F: 5'-GGC ACG GCT GTC CAA GGA G-3' SEQ ID. NO.41; reverse, 15 apoE-R: 5'-AGG CCG CGC TCG GCG CCC TC-3' SEQ ID. NO.42). Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs from Pharmacia (Freiburg, Germany). The total reaction volume was 50 μ L including 8 pmol of each primer and 10% DMSO (dimethylsulfoxide, Sigma) with approximately 200 ng of genomic 20 DNA used as template. Solutions were heated to 80°C before the addition of 1U polymerase; PCR conditions were: 2 min at 94°C, followed by 40 cycles of 30 sec at 94°C, 45 sec at 63°C, 30 sec at 72°C, and a final extension time of 2 min at 72°C.

Restriction enzyme digestion and polyacrylamide electrophoresis.

25 CfoI and RsaI and reaction buffer L were purchased from Boehringer-Mannheim, and HhaI from Pharmacia (Freiburg, Germany). For CfoI alone and simultaneous CfoI/RsaI digestion, 20 pL of amplified products were diluted with 15 μ L water and 4 pL Boehringer-Mannheim buffer L; after addition of 10 units of appropriate restriction enzyme(s)

-136-

the samples were incubated for 60 min at 37°C. The procedure for simultaneous HhaI/RsaI digestion required first digestion by RsaI in buffer L for one hour followed by addition of NaCl (50 mM end concentration) and HhaI, and additional incubation for one hour. 20 μ L of the restriction
5 digest were analyzed on a 12% polyacrylamide gel as described elsewhere (Hixson (1990) J. Lipid Res. 31:545-548). Recognition sequences of RsaI and CfoI (HhaI) are GT/AC and GCG/C, respectively; masses of expected digestion fragments from the 252-mer amplified product with CfoI alone and the simultaneous double digest with CfoI (or
10 HhaI) and RsaI are given in Table V.

Thermo-PROBE.

PCR amplification was performed as described above, but with products purified with the Qiagen' Qiaquick' kit to remove unincorporated primers. Multiplex Thermo-PROBE was performed with
15 35 μ L amplified product and 8 pmol each of the codon 112 (5'-GCG GAC ATG GAG GAC GTG-3' SEQ ID. NO.43) and 158 (5'-GAT GCC GAT GAC CTG CAG AAG-3' SEQ ID. NO.44) detection primers in 20 μ L including ~1 pmol purified biotinylated antisense template immobilized on streptavidin coated magnetic beads, 2.5 units Thermosequenase,
20 2 μ L Thermosequenase buffer, 50 μ M of each dNTP and 200 μ M of ddXTP, with the base identity of N and X as described in the text. Cycling conditions were: denaturation (94°C, 30 sec) followed by 30 cycles at 94°C (10 min) and 60°C (45 sec).

Sample preparation and analysis by MALDI-TOF MS.

25 For precipitation (Stults *et al.*, (1991) Rapid Commun. Mass Spectrom. 5: 359-363) of both digests and PROBE products, 5 μ L 3M ammonium acetate (pH 6.5), 0.5 μ L glycogen (10 mg/ml, Sigma), and 110 μ L absolute ethanol were added to 50 μ L of the analyte solutions and stored for 1 hour at room temperature. After 10 min centrifugation at

-137-

13,000 X g the pellet was washed in 70% ethanol and resuspended in 1 μ l 18 Mohm/cm H₂O water. Where noted in the text, additional desalting was achieved by shaking 10-20 μ L of ammonium saturated DOWEX (Fluka #44485) cation exchange beads in 40 μ L of analyte. The beads, 5 purchased in the protonated form, were pre-treated with three 5 min spin-decant steps in 2M NH₄OH, followed with H₂O and 10 mM ammonium citrate.

0.35 μ L of resuspended DNA was mixed with 0.35-1.3 μ L matrix solutions (Wu *et al.* (1993) Rapid Commun. Mass Spectrom. 7:142-146) 10 0.7 M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium citrate in 1:1 H₂O:CH₃CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular 15 masses ($M_r(\text{calc})$) of the fragments were calculated from atomic compositions; the mass of a proton (1.08 Da) is subtracted from raw data values in reporting experimental molecular masses ($M_r(\text{exp})$) as neutral basis. An external calibration generated from eight peaks (3000-18000 Da) was applied to all spectra.

20 RESULTS

Digestion with Cfol alone.

The inset to Figure 57a shows a 12% polyacrylamide gel electrophoretic separation of an $\epsilon 3/\epsilon 3$ genotype after digestion of the 252 bp apo E amplified product with Cfol. Comparison of the 25 electrophoretic bands with a molecular weight ladder shows the cutting pattern to be as mostly as expected (Table V) for the $\epsilon 3/\epsilon 3$ genotype. Differences are that the faint band at approximately 25 bp is not expected, and the smallest fragments are not observed. The accompanying mass spectrum of precipitated digest products shows a

-138-

similar pattern, albeit at higher resolution. Comparison with Table V shows that the observed masses are consistent with those of single-stranded DNA; the combination of an acidic matrix environment (3-HPA, pK_a 3) and the absorption of thermal energy via interactions with the 337

- 5 nm absorbing 3-HPA upon ionization is known to denature short stretches of dsDNA under normal MALDI conditions (Tang, K et al., (1994) Rapid Commun Mass Spectrom 8:183-186).

- The approximately 25-mers, unresolved with electrophoresis, are resolved by MS as three single stranded fragments; while the largest
- 10 (7427 Da) of these may represent a doubly charged ion from the 14.8 kDa fragments ($m = 14850$, $z = 2$; $m/z = 7425$), the 6715 and 7153 Da fragments could result from PCR artifacts or primer impurities; all three peaks are not observed when amplified products are purified with Qiagen purification kits prior to digestion. The Table V 8871 Da 29-mer
- 15 sense strand 3'-terminal fragment is not observed; the species detected at 9186 Da is consistent with the addition of an extra base (9187 - 8871 = 316, consistent with A) by the Taq-polymerase during PCR amplification (Hu, G et al., (1993) DNA and Cell Biol 12:763-770). The individual single strands of each double strand with <35 bases (11 kDa)
- 20 are resolved as single peaks; the 48-base single strands ($M_r(\text{calc})$ 14845 and 14858), however, are observed as an unresolved single peak at 14850 Da. Separating these into single peaks would require a mass resolution ($m/\Delta m$, the ratio of the mass to the peak width at half height) of $14850/13 = 1140$, nearly an order of magnitude greater than what is
- 25 routine with the standard reflectron time-of-flight instrumentation used in this study; resolving such small mass differences with high performance instrumentation such as Fourier transform MS, which provides up to three orders of magnitude higher resolution in this mass range, has been demonstrated. The 91-mer single strands ($M_r(\text{calc})$ 27849 and 28436)

-139-

are also not resolved, even though this requires a resolution of only < 50 . The dramatic decrease in peak quality at higher masses is due to metastable fragmentation (i.e. depurination) resulting from excess internal energy absorbed during and subsequent to laser irradiation.

5 *Simultaneous digestion with Cfol and RsaI.*

Figure 57b (inset) shows a 12% polyacrylamide gel electrophoresis separation of $\epsilon 3/\epsilon 3$ double digest products, with bands consistent with dsDNA with 24, 31, 36, 48, and 55 base pairs, but not for the smaller fragments. Although more peaks are generated (Table V) than with Cfol alone, the corresponding mass spectrum is more easily interpreted and reproducible since all fragments contain < 60 bases, a size range far more appropriate for MALDI-MS if reasonably accurate M_r values (e.g., 0.1%) are desired. For fragments in this mass range, the mass measuring accuracy using external calibration is -0.1% (i.e. $\leq +10$ Da at 10 kDa). Significant depurination (indicated in Figure by asterisk) is observed for all peaks above 10 kDa, but even the largest peak at 17171 Da is clearly resolved from its depurination peak so that an accurate M_r can be measured. Although molar concentrations of digest products should be identical, some discrimination against those fragments with ≤ 11 bases is observed, probably due to their loss in the ethanol/glycogen precipitation step. The quality of MS results from simultaneous digestion with Cfol (or HhaI) and RsaI is superior to those with Cfol (or HhaI) alone, since the smaller fragments generated are good for higher mass accuracy measurements, and with all genotypes there is no possibility for dimer peaks overlapping with high mass diagnostic peaks. Since digestion by RsaI/Cfol and RsaI/HhaI produce the same restriction fragments but the former may be performed as a simultaneous digest since their buffer requirements are the same, this enzyme mixture was used for all subsequent genotyping by restriction digest protocols.

-140-

Table V

Mass and Copy Number of Expected Restriction Digest Products

Table Va Cfol Digestion^a

	(+) (-)	e2/e2	e2/e2	e2/e2	e2/e2	e2/e2	e2/e2
5	5781, 5999	--	--	1	--	1	2
	10752, 10921	--	1	1	2	2	2
	14845, 14858	--	1	1	2	2	2
	22102, 22440	--	--	1	--	1	2
	25575, 25763	2	1	1	--	--	--
10	27849, 28436	2	2	1	2	1	--

Table Vb. Cfol/Rsal Digestion^b

	(+) (-)	e2/e2	e2/e3	e2/e4	e3/e3	e3/e4	e4/e4
	3428, 4025	--	1	1	2	2	2
15	5283, 5880	--	--	1	--	1	2
	5781, 5999	--	--	1	--	1	2
	11279, 11627	2	2	1	2	1	--
	14845, 14858	--	1	1	2	2	2
	18269, 18848	2	2	1	--	--	--

20

^aCfol Invariant fragment masses: 1848, 2177, 2186, 2435, 4924, 5004, 5412, 5750, 8871, 9628 Da.

-141-

^bCfoI/RsaI Invariant fragment masses: 1848, 2177, 2186, 2436, 4924, 5004, 5412, 5750, 6745, 7510, 8871, 9628, 16240, 17175 Da.

Table VI					
	ddT M _r (Calc)	ddT M _r (Exp)	ddC M _s (Calc)	ddC M _r (Exp)	
5	e2/e2	^a 5918, ^b 6768	_____	^a 6536, ^b 7387	_____
	e2/e3	^a 5918, ^b 6768, ^b 7965	5919, 6769, 7967	^a 6536, ^b 6753, ^b 7387	6542, 6752, 7393
	e2/e4	^a 5918, ^b 6768, ^b 7965, ^a 8970	_____	^a 5903, ^b 6536, ^b 6753, ^a 7387	_____
	e3/e3	^a 5918, ^b 7965	5918, 7966	^a 6536, ^b 6753	6542, 6756
	e3/e4	^a 5918, ^b 7965, ^a 8970	5914, 7959, 8965	^a 5903, ^b 6536, ^b 6753	5898, 6533, 6747
10	e4/e4	^b 7965, ^a 8970	7966, 8969	^a 5903, ^b 6753	5900, 6752

^aFrom codon 112 detection primer (unextended 5629.7 Da).

^bFrom codon 158 detection primer (unextended 6480.3 Da).

Dashed lines: this genotype not available from the analyzed pool of 100 patients.

- 15 Figure 58a-c shows the ApoE $\epsilon 3/\epsilon 3$ genotype after digestion with CfoI and a variety of precipitation schemes; equal volume aliquots of the same amplified product were used for each. The sample treated with a single precipitation (Figure 58a) from an ammonium acetate/ethanol/glycogen solution results in a mass spectrum characterized by broad peaks, especially at high mass. The masses for intense peaks at 5.4, 10.7, and 14.9 kDa are 26 Da (0.5%), 61 Da (0.6%), and 45 Da (0.3%) Da higher, respectively, than the expected values; the resolution (the ratio of a peak width at half its total intensity to the measured mass of the peak) for each of these is ~50, and decreases with increasing mass. Such
- 20

-142-

observations are consistent with a high level of nonvolatile cation adduction; for the 10.8 kDa fragment, the observed mass shift is consistent with a greater than unit ratio of adducted:nonadducted molecular ions.

- 5 MS peaks from a sample redissolved and precipitated a second time are far sharper (Figure 58b), with resolution values nearly double those of the corresponding Figure 58a peaks. Mass accuracy values are also considerably improved; each is within 0.07% of its respective calculated values, close to the independently determined instrumental
- 10 limits for DNA measurement using 3-HPA as a matrix. Single (not shown) and double (Figure 58C) precipitations with isopropyl alcohol (IPA) instead of ethanol result in resolution and mass accuracy values comparable to those for corresponding ethanol precipitations, but enhanced levels of dimerization are observed, again potentially confusing
- 15 measurements when such dimers overlap with higher mass "diagnostics" monomers present in the solution. EtOH/ammonium acetate precipitation with glycogen as a nucleation agent results in nearly quantitative recovery of fragments except for the 7-mers, serving as a simultaneous concentration and desalting step prior to MS detection. Precipitation
- 20 from the same EtOH/ammonium acetate solutions in the absence of glycogen results in far poorer recovery, especially at low mass.

The results indicate that to obtain accurate ($M_r(\text{exp})$) values after either 1PA and EtOH precipitations, a second precipitation is necessary to maintain high mass accuracy and resolution.

- 25 The ratio of matrix:digest product also affects spectral quality; severe suppression of higher mass fragments (not shown) observed with 1:1 volume matrix: digest product (redissolved in $1\mu\text{L}$) is alleviated by using a 3 - 5 fold volume excess of matrix.

-143-

Apo E genotyping by enzymatic digestion. Codon 112 and 158 polymorphisms fall within CfoI (but not RsaI) recognition sequences. In the 252 bp amplified product studied here, invariant (i.e. cut in all genotypes) sites cause cuts after bases 31, 47, 138, 156, 239, and

5 246. The cutting site after base 66 is only present for $\epsilon 4$, while that after base 204 is present in $\epsilon 3$ and $\epsilon 4$; the $\epsilon 2$ genotype is cut at neither of these sites. These differences in the restriction pattern can be demonstrated as variations in mass spectra. Figure 59 shows mass spectra from several ApoE genotypes available from a pool of 100

10 patients (Braun, A *et al.*, (1992) Hum. Genet. 89:401-406). Vertical dashed lines are drawn through those masses corresponding to the expected Table V diagnostic fragments; other labeled fragments are invariant. Referring to Table V, note that a fragment is only considered "invariant" if it is present in duplicate copies for a given allele; to satisfy

15 this requirement, such a fragment must be generated in each of the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ alleles.

The spectrum in Figure 59a contains all of the expected invariant fragments above 3 kDa, as well as diagnostic peaks at 3428 and 4021 (both weak), 11276 and 11627 (both intense), 14845, 18271, and

20 18865 Da. The spectrum in Figure 59b is nearly identical except that the pair of peaks at 18 kDa is not detected, and the relative peak intensities, most notably among the 11-18 kDa fragments, are different. The spectrum in Figure 59c also has no 18 kDa fragments, but instead has new low intensity peaks between 5-6 kDa. The intensity ratios for

25 fragments above 9 kDa are similar to those of Figure 59b except for a relatively lower 11 kDa fragment pair. Figure 59d, which again contains the 5-6 kDa cluster of peaks, is the only spectrum with no 11 kDa fragments, and like the previous two also has no 18 kDa fragment.

-144-

Despite the myriad of peaks in each spectrum, each genotype can be identified by the presence and absence of only a few of the Table Vb diagnostic peaks. Due to the limited resolution of the MALDI-TOF instrumentation employed, the most difficult genotypes to differentiate

5 are those based upon the presence or absence of the four diagnostic fragments between 5.2 and 6.0 kDa characteristic of the $\epsilon 4$ allele, since these fragments nearly overlap with several invariant peaks. It has been found herein that the 5283 Da diagnostic fragment overlaps with a depurination peak from the 5412 Da invariant fragment, and the 5781 Da

10 diagnostic peak is normally not completely resolved from the 5750 Da invariant fragment. Thus, distinguishing between an $\epsilon 2/\epsilon 4$ and $\epsilon 2/\epsilon 3$, or between an $\epsilon 3/\epsilon 4$ and an $\epsilon 3/\epsilon 3$ allele, relies upon the presence or absence of the 5880 and 5999 Da fragments. Each of these is present in Figures 59c and 59d, but not in 59a or 59b.

15 The genotype of each of the patients in Figure 59 can be more rapidly identified by reference to the flowchart in Figure 60. Consider the Figure 59a spectrum. The intense pair of peaks at 11 kDa discounts the possibility of homozygous $\epsilon 4$, but does not differentiate between the other five genotypes. Likewise, the presence of the unresolved 14.8 kDa

20 fragments is inconsistent with homozygous $\epsilon 2$, but leaves four possibilities ($\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 3$, $\epsilon 3/\epsilon 4$). Of these only $\epsilon 2/\epsilon 3$ and $\epsilon 2/\epsilon 4$ are consistent with the 18 kDa peaks; the lack of peaks at 5283, 5879, 5779, and 5998, Da indicate that the Figure 59a sample is $\epsilon 2/\epsilon 3$. Using the same procedure, the Figures 59b-d genotypes can be identified as

25 $\epsilon 3/\epsilon 3$, $\epsilon 3/\epsilon 4$, and $\epsilon 4/\epsilon 4$, respectively. To date, all allele identifications by this method have been consistent with, and in many cases more easily interpreted than, those attained via conventional methods. The assignment can be further confirmed by assuring that fragment intensity ratios are consistent with the copy numbers of Table V. For instance,

-145-

the 14.8 kDa fragments are of lower intensity than those at 16- 17 kDa in Figure 59a, but the opposite is seen in Figures 59b-d. This is as expected, since in the latter three genotypes the 14.8 kDa fragments are present in duplicate, but the first is a heterozygote containing $\epsilon 2$, so that half of the amplified products do not contribute to the 14.8 kDa signal. Likewise, comparison of the 11 kDa fragment intensity to those at 9.6 and 14.8 kDa indicate that this fragment is double, double, single, and zero copy in Figures 59a, d, respectively. These data confirm that MALDI can perform in a semi-quantitative way under these conditions.

- 10 *ApoE genotyping by Primer Oligo Base Extension (PROBE)*. The PROBE reaction was also tested as a means of simultaneous detection of the codon 112 and 158 polymorphisms. A detection primer is annealed to a single-stranded PCR-amplified template so that its 3' terminus is just downstream of the variable site. Extension of this primer by a DNA
15 polymerase in the presence of three dNTPs and one ddXTP (that is not present as a dNTP) results in products whose length and mass depend upon the identity of the polymorphic base. Unlike standard Sanger type sequencing, in which a particular base-specific tube contains -99% dXTP and -1% ddXTP, the PROBE mixture contains 100% of a particular
20 ddXTP combined with the other three dNTPs. Thus with PROBE a full stop of all detection primers is achieved after the first base complementary to the ddXTP is reached.

For the $\epsilon 2/\epsilon 3$ genotype, the PROBE reaction (mixture of ddTTP, dATP, dCTP, dGTP) causes a $M_r(\text{exp})$ shift of the codon 112 primer to
25 5919 Da, and of the codon 158 primer to 6769 and 7967 Da (Table VI); a pair of extension products results from the single codon 158 primer because the $\epsilon 2/\epsilon 3$ genotype is heterozygous at this position. Three extension products (one from codon 158, two from 112) are also observed from the heterozygote $\epsilon 3/\epsilon 4$ (Figure 61c and Table VI), while

-146-

only two products (one from each primer) are observed from the Figure 61b ($\epsilon 3/\epsilon 3$) and Figure 59d ($\epsilon 4/\epsilon 4$) homozygote alleles. Referring to Table VI, each of the available alleles result in all expected ddT reaction product masses within 0.1% of the theoretical mass, and thus each is

5 unambiguously characterized by this data alone. Further configuration of the allele identities may be obtained by repeating the reaction with ddCTP (plus dATP, dTTP, dGTP); these results, summarized also in Table VI, unambiguously confirm the ddT results.

Appropriateness of the methods. Comparison of Figures 59

10 (restriction digestion) and 61 (PROBE) indicates that the PROBE method provides far more easily interpreted spectra for the multiplex analysis of codon 112 and 158 polymorphisms than does the restriction digest analysis. While the digests generate up to -25 peaks per mass spectrum and in some case diagnostic fragments overlapping with invariant

15 fragments, the PROBE reaction generates a maximum of only two peaks per detection primer (i.e. polymorphism). Automated peak detection, spectrum analysis, and allele identification would clearly be far more straightforward for the latter. Spectra for highly multiplexed PROBE, in which several polymorphic sites from the same or different amplified

20 products are measured from one tube, are also potentially simple to analyze. Underscoring its flexibility, PROBE data analysis can be further simplified by judicious a priori choice of primer lengths, which can be designed so that no primers or products can overlap in mass.

Thus while PROBE is the method of choice for large scale clinical

25 testing of previously well characterized polymorphic sites, the restriction digest analysis as described here is ideally suited to screening for new mutations. The identity of each of the two polymorphisms discussed in this study affects the fragment pattern; if this is the only information used, then the MS detection is a faster alternative to conventional

-147-

- electrophoretic separation of restriction fragment length polymorphism products. The exact measurement of fragment M_r values can also give information on about sites completely remote from the enzyme recognition site since other single point mutations necessarily alter the
- 5 mass of each of the single strands of the double stranded fragment containing the mutation. The 252 bp amplified product could also contain allelic variants resulting in, for example, previously described Gly127 Asp (Weisgraber, KH *et al.*, (1984) J. Clin. Invest. **73**:1024-1033), Arg136Ser (Wardell, MR *et al.*, (1987) J. Clin. Invest. **80**:483-10 490), Arg142Cys (Horie, Y *et al.*, (1992) J. Biol. Chem. **267**:1962-1968), Arg145Cys (Rall SC Jr *et al.*, (1982) Proc. Natl. Acad. Sci. U.S.A. **79**:4696-4700), Lys146Glu (Mann, WA *et al.*, (1995) J. Clin. Invest. **96**:1100-1107), or Lys146Gln (Smit, M *et al.*, (1990) J. Lipid Res. **31**:45-53) substitutions. The G→A base substitution which codes for the
- 15 Gly127 Asp amino acid substitution would result in a -16 Da shift in the sense strand, and in a +15 Da (C→T) shift in the antisense strand, but not in a change in the restriction pattern. Such a minor change would be virtually invisible by electrophoresis; however, with accurate mass determination the substitution could be detected; the invariant 55-mer
- 20 fragment at 16240 (sense) and 17175 Da would shift to 16224 and 17190 Da, respectively. Obtaining the mass accuracy required to detect such minor mass shifts using current MALDI-TOF instrumentation, even with internal calibration, is not routine since minor unresolved adducts and/or poorly defined peaks limit the ability for accurate mass calling.
- 25 With high performance electrospray ionization Fourier transform (ESI-FTMS) single Da accuracy has been achieved with synthetic oligonucleotides (Little, DP *et al.*, (1995) Proc. Natl. Acad. Sci. U.S.A. **92**:2318-2322) up to 100-mers (Little, DP *et al.*, (1994) J. Am. Chem. Soc. **116**:4893-4897), and similar results have recently been achieved

-148-

with up to 25-mers using MALDI-FTMS (Li, Y *et al.*, (1996) Anal. Chem. 68:2090-2096).

EXAMPLE 13

A Method for Mass Spectrometric Detection of DNA Fragments Associated With Telomerase Activity

INTRODUCTION

One-fourth of all deaths in the United States are due to malignant tumors (R.K. Jain, (1996) Science 271:1079-1080). For diagnostic and therapeutic purposes there is a high interest in reliable and sensitive methods of tumor cell detection.

Malignant cells can be distinguished from normal cells by different properties. One of those is the immortalization of malignant cells which enables uncontrolled cell-proliferation. Normal diploid mammalian cells undergo a finite number of population doublings in culture, before they undergo senescence. It is supposed that the number of population doublings in culture, before they undergo senescence. It is supposed that the number of population doublings is related to the shortening of chromosome ends, called telomers, in every cell division. The reason for said shortening is based on the properties of the conventional semiconservative replication machinery. DNA polymerases only work in 5' to 3' direction and need an RNA primer.

Immortalization is thought to be associated with the expression of active telomerase. Said telomerase is a ribonucleoprotein catalyzing repetitive elongation of templates. This activity can be detected in a native protein extract of telomerase containing cells by a special PCR-system (N.W. Kim *et al.* (1994) Science 266:2011-2015) known as telomeric repeat amplification protocol (TRAP). The assay, as used herein, is based on the telomerase specific extension of a substrate primer (TS) and a subsequent amplification of the telomerase specific

-149-

extension products by a PCR step using a second primer (bioCX) complementary to the repeat structure. The characteristic ladder fragments of those assays are conventionally detected by the use of gel electrophoretic and labeling or staining systems. These methods can be replaced by MALDI-TOF mass spectrometry leading to faster accurate and automated detection.

MATERIALS AND METHODS

Preparation of cells

1 x 10⁶ cultured telomerase-positive cells were pelleted, washed once with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄•7H₂O, 1.4 mM KH₂PO₄ in sterile DEPC water). The prepared cells may be stored at -75°C. Tissue samples have to be homogenized, according to procedures well known in the art, before extraction.

Telomerase extraction

Pellet was resuspended in 200 µl CHAPS lysis buffer (10 mM Tris-HCl pH 7.5, 1 mM MgCl₂, 1 mM EGTA, 0.1 mM benzamidine, 5 mM β-mercaptoethanol, 0.5% CHAPS, 10% glycerol) and incubated on ice for 30 min. The sample was centrifuged at 12,000 g for 30 min at 4°C. The supernatant was transferred into a fresh tube and stored at 75°C until use.

TRAP-assay

2 µl of telomerase extract were added to a mixture of 10x TRAP buffer (200 mM Tris-HCl pH 8.3, 15 mM MgCl₂, 630 mM KCl, 0.05% Tween 20, 10 mM EGTA) 50x dNTP-mix (2.5 mM each dATP, dTTP, dGTP, and dCTP), 10 pmol of TS primer and 50 pmol of bio CX primer in a final volume of 50 µl. The mixture was incubated at 30°C for 10 minutes and 5 min. at 94°C, 2 units of Taq Polymerase were added and a PCR was performed with 30 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 45 seconds.

-150-

Purification of TRAP-assay products

- For every TRAP-assay to be purified, 50 μ l Streptavidin M-280 Dynabeads (10 mg/ml) were washed twice with 1x BW buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl). 50 μ l of 2x BW buffer were
- 5 added to the PCR mix and the beads were resuspended in this mixture. The beads were incubated under gentle shaking for 15 min. at ambient temperature. The supernatant was removed and the beads were washed twice with 1x BW buffer. To the beads 50 μ l 25% ammonium hydroxide were added and incubated at 60°C for 10 min. The supernatant was
- 10 saved, the procedure repeated, both supernatants were pooled and 300 μ l ethanol (100%) were added. After 30 min. the DNA was pelleted at 13,000 rpm for 12 min., the pellet was air-dried and resuspended in 600 nl ultrapure water.

MALDI-TOF MS of TRAP-assay products

- 15 300 nl sample were mixed with 500 nl of saturated matrix-solution (3-HPA:ammonium citrate = 10:1 molar ratio in 50% aqueous acetonitrile), dried at ambient temperature and introduced into the mass spectrometer (Vision 2000, Finigan MAT). All spectra were collected in reflector mode using external calibration.

20 *Sequences and masses*

bioCX: d(bio-CCC TTA CCC TTA CCC TTA CCC TAA SEQ ID NO. 45),
mass: 7540 Da.

TS: d(AAT CCG TGC AGC AGA GTT SEQ ID NO.46), mass: 5523 Da.

Telomeric-repeat structure: (TTAGGG)_n, mass of one repeat: 1909.2

25 *Amplification products:*

TS elongated by three telomeric repeats (first amplification product):
12452 Da. (N₃)

TS elongated by four telomeric repeats: 14361 Da. (N₄)

TS elongated by seven telomeric repeats: 20088 Da. (N₇)

-151-

RESULTS

Figure 62 depicts a section of a TRAP-assay MALDI-TOF mass spectrum. Assigned are the primers TS and bioCX at 5497 and 7537 Da, respectively (calculated 5523 and 7540 Da). The signal marked by an asterisk represents n-1 primer product of chemical DNA synthesis. The first telomerase specific TRAP-assay product is assigned at 12775 Da. This product represents a 40-mer containing three telomeric repeats. Due to primer sequences this is the first expected amplification product of a positive TRAP-assay. The product is elongated by an additional nucleotide due to extendase activity of Taq DNA polymerase (calculated non-extended product: 12452 Da, by A extended product: 12765 Da). The signal at 6389 Da represents the doubly charged ion of this product (calculated: 6387 Da). Figure 63 shows a section of higher masses of the same spectrum as depicted in figure 62, therefore the signal at 12775 Da is identical to that in figure 62. The TRAP-assay product containing seven telomeric repeats, representing a 64-mer also elongated by an additional nucleotide, is detected at 20322 Da (calculated: 20395 Da). The signals marked 1, 2, 3 and 4 cannot be base-line resolved. This region includes of: 1. signal of dimeric n-1 primer, 2. second TRAP-assay amplification product, containing 4 telemeric repeats and therefore representing a 46-mer (calculated: 14341 Da/14674 Da for extendase elongated product) and 3. dimeric primer-ion and furthermore all their corresponding depurination signals. There is a gap observed between the signals of the second and fifth extension product. This signal gap corresponds to the reduced band intensities observed in some cases for the third and fourth extension product in autoradiographic analysis of TRAP-assays (N.W. Kim *et al.* (1994) *Science* 266:2013).

The above-mentioned problems, caused by the dimeric primer and related signals, can be overcome using an ultrafiltration step employing a

-152-

molecular weight cut-off membrane for primer removal prior to MALDI-TOF-MS analysis. This will permit an unambiguous assignment of the second amplification product.

EXAMPLE 14

5 A method for Detecting Neuroblastoma-Specific Nested RT-amplified products Via MALDI-TOF Mass Spectrometry

Introduction

Neuroblastoma is predominantly a tumor of early childhood with 66% of the cases presenting in children younger than 5 years of age.

- 10 The most common symptoms are those due to tumor mass, bone pain, or those caused by excessive catecholamine secretion. In rare cases, neuroblastoma can be identified prenatally (R.W. Jennings et al, (1993) J. Ped. Surgery 28:1168-1174). Approximately 70% of all patients with neuroblastoma have metastatic disease at diagnosis. The prognosis is
- 15 dependent on age at diagnosis, clinical stage and other parameters.

For diagnostic purposes there is a high interest in reliable and sensitive methods of tumor cell detection, e.g., in control of autologous bone marrow transplants or on-going therapy.

- Since catecholamine synthesis is a characteristic property of
- 20 neuroblastoma cells and bone marrow cells lack this activity (H. Naito et al., (1991) Eur. J. Cancer 27:762-765), neuroblastoma cells or metastasis in bone marrow can be identified by detection of human tyrosine 3-hydroxylase (E.C. 1.14.16.2, hTH) which catalyzes the first step in biosynthesis of catecholamines.

- 25 The expression of hTH can be detected via reverse transcription (RT) polymerase chain reaction (PCR) and the amplified product can be analyzed via MALDI-TOF mass spectrometry.

Materials and methods

Cell- or tissue-treatment

-153-

Cultures cells were pelleted (10 min. 8000 rpm) and washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄•7H₂O, 1.4 mM KH₂PO₄ in sterile PEPC water). The pellet was resuspended in 1 ml lysis/binding buffer (100 mM Tris-HCl, pH 8.0, 500 mM LiCl, 10 mM

5 EDTA, 1% Li-dodecyl sulfate, 5 mM DTT) until the solution becomes viscose. Viscosity was reduced by DNA-shear step using a 1 ml syringe. The lysate may be stored in -75°C or processed further directly. Solid tissues (e.g., patient samples) have to be homogenized before lysis.

Preparation of magnetic Oligo-dT(25) beads

- 10 100 µL beads per 1x10⁶ cells were separated from the storage buffer and washed twice with 200 µL lysis/binding buffer.

Isolation of poly A⁺ RNA

- The cell lysate was added to the prepared beads and incubated for 5 min. at ambient temperature. The beads were separated magnetically
- 15 for 2-5 min. and washed twice with 0.5 ml LDS (10 mM Tris-HCl, pH 8.0, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS).

Solid-phase first-strand cDNA synthesis

- The poly A⁺RNA containing beads were resuspended in 20 µL of reverse transcription mix (50 mM Tris-HCl, pH 8.3, 8 mM MgCl₂, 30 mM
- 20 KCl, 10 mM DTT, 1.7 mM dNTPs, 3 U AMV reverse transcriptase) and incubated for 1 hour at 45°C (with a resuspension step all ten min.). The beads were separated from the reverse transcription mix, resuspended in 50 µL of elution buffer (2 mM EDTA pH 8.0) and heated to 95°C for 1 min. for elution of the RNA. The beads with the cDNA
- 25 first-strand can be stored in TB (0.089 M Tris-base, 0.089 M boric acid, 0.2 mM EDTA pH 8.0), TE 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) or 70% ethanol for further processing.

Nested polymerase chain reaction

-154-

Beads containing cDNA first-strand were washed twice with 1x PCR buffer (20 mM Tris-HCl pH 8.75, 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgSO_4 , 0.1% Triton X-100, 0.1 mg bovine serum albumin) and resuspended in PCR mix (containing 100) pmol of each outer primer, 2.5

5 u *Pfu* (exo-) DNA polymerase, 200 μM of each dNTP and PCR buffer in a final volume of 50 μL). The mixture was incubated at 72°C 1 min. and amplified by PCR for 30 cycles. for the nested reaction: 1 μL of the first PCR was added as template to a PCR mix d(as above but nested primers instead of outer primers) and subjected to the following temperature

10 program: 94°C 1 min., 65°C 1 min. and 72°C 1 min. for 20 cycles.

Purification of nested amplified products

Primers and low-molecular reaction by-products are removed using 10,000 Da cut-off ultrafiltration-unit. Ultrafiltration was performed at 7,500 g for 25 minutes. For every PCR to be purified, 50 μL Streptavidin

15 M-280 Dynabeads (10 mg/ml) were washed twice with 1xBW buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl), added to the ultrafiltration membrane and incubated under gentle shaking for 15 min. at ambient temperature. The supernatant was removed and the beads were washed twice with 1xBW buffer. 50 μL 25% ammonium hydroxide

20 were added to the beads and incubated at ambient temperature for 10 min. The supernatant was saved, the procedure repeated, both supernatants were pooled and 300 μL ethanol (100%) were added. After 30 min. the DNA was pelleted at 13,000 rpm for 12 min., the pellet was air-dried and resuspended in 600 nl ultrapure water.

25 *MALDI-TOF MS of nested amplified products*

300 nl sample was mixed with 500 nl of saturated matrix-solution (3-HPA: ammonium citrate = 10:1 molar ratio in 50% aqueous acetonitrile), dried at ambient temperature and introduced into the mass

-155-

spectrometer (Vision 2000, Finigan MAT). All spectra were collected in reflector mode using external calibration.

Outer primers:

hTH1: d(TGT CAG AGC TGG ACA AGT GT SEQ ID NO:47)

5 hTH2: d(GAT ATT GTC TTC CCG GTA GC SEQ ID NO:48)

Nested primers:

bio-hTH d(bio-CTC GGA CCA GGT GTA CCG CC SEQ ID NO:49),

mass: 6485 Da

hTH6; d(CCT GTA CTG GAA GGC GAT CTC SEQ ID NO:50),

10 mass:6422 21 Da

mass of biotinylated single strand amplified product: 19253.6 Da

mass of nonbiotinylated single strand amplified product: 18758.2 Da

Results

15 A MALDI-TOF mass spectrum of a human tyrosine 3-hydroxylase (hTH) specific nested amplified product (61-mer) is depicted in figure 64. The signal at 18763 Da corresponds to non-biotinylated strand of the amplified product (calculated: 18758.2 Da, mass error: 0.02 Da). The signals below 10,000 and above 35,000 Da are due to multiply charged and dimeric amplified product-ions, respectively.

20 The product was obtained from a solid phase cDNA derived in a reverse transcription reaction from 1×10^6 cells of a neuroblastoma cell-line (L-A-N-1) as described above. The cDNA first-strand was subjected to a first PCR using outer primers (hTH1 and hTH2), an aliquot of this

25 PCR was used as template in a second PCR using nested primers (biohTH and hTH6). The nested amplified product was purified and MALDI-TOF MS analyzed:

-156-

The spectrum in Fig. 64 demonstrates the possibility of neuroblastoma cell detection using nested RT-PCR and MALDI-TOF MS analysis.

EXAMPLE 15

5 Rapid Detection of the RET Proto-oncogene Codon 634 Mutation Using Mass Spectrometry

Material and Methods

Probe

- The identity of codon 634 in each of the three alleles was
- 10 confirmed by RsaI enzymatic digestion, single strand conformational polymorphism or Sanger sequencing. Exon 11 of the RET gene was PCR amplified (40 cycles) from genomic DNA using Taq-Polymerase (Boehringer-Mannheim) with 8 pmol each of 5'-biotinylated forward (5'-biotin-CAT GAG GCA GAG CAT ACG CA-3' SEQ ID NO:51) and
- 15 unmodified reverse (5'-GAC AGC AGC ACC GAG ACG AT-3' SEQ ID NO:52) primer per tube; amplified products were purified using the Qiagen (QIAquick" kit to remove unincorporated primers. 15 μ L of amplified product were immobilized on 10 μ L (10 mg/mL) Dynal streptavidin coated magnetic beads, denatured using the manufacturer's
- 20 protocol, and the supernatant containing antisense strand discarded, the PROBE reaction was performed using thermoSequenase (TS) DNA Polymerase (Amersham) and Pharmacia dNTP/ddNTPs. 8 pmol of extension primer (5'-CGG CTG CGA TCA CCG TGC GG-3' SEQ ID NO:53) was added to 13 μ L H₂O, 2 μ L TS-buffer, 2 μ L 2mM ddATP (or
- 25 ddTTP), and 2 μ L of 0.5 mM dGTP/dCTP/dTTP (or dGTP/DCTP/dATP), and the mixture heated for 30 sec @ 94°C, followed by 30 cycles of 10 sec @ 94°C and 45 sec @ 50°C; after a 5 min. incubation @ 95°C, the supernatant was decanted, and products were desalted by ethanol precipitation with the addition of 0.5 μ L of 10mg/mL glycogen. The

-157-

resulting pellet was washed in 70% ethanol, air dried, and suspended in 1 μ L H₂O. 300 nL of this was mixed with the MALDI matrix (0.7 M 3-hydroxypicolinic acid, 0.07 M ammonium citrate in 1:1 H₂O:CH₃CN) on a stainless steel sample probe and air dried. Mass spectra were collected on a Thermo Bionalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Experimental masses ($m_r(\text{exp})$) reported are those of the neutral molecules as measured using external calibration.

Direct Measurement of Diagnostic Products

- 10 PCR amplifications conditions for a 44 bp region containing codon 634 were the same as above but using Pfu polymerase; the forward primer contained a ribonucleotide at its 3'-terminus (forward, 5'-GAT CCA CTG TGC GAC GAG C (SEQ ID NO:54) -ribo; reverse, 5'-GCG GCT GCG ATC ACC GTG C (SEQ ID NO:55). After product immobilization
- 15 and washing, 80 μ L of 12.5% NH₄OH was added and heated at 80°C overnight to cleave the primer from 44-mer (sense strand) to give a 25-mer. Supernatant was pipetted off while still hot, dried resuspended in 50 μ L H₂O, precipitated, resuspended, and measured by MALDI-TOF as above. MALDI-FTMS spectra of 25-mer synthetic analogs were collected
- 20 as previously described (Li, Y. *et al.*, (1996) *Anal. Chem.* **68**:2090-2096); briefly, 1-10 pmol DNA was mixed 1:1 with matrix on a direct insertion probe, admitted into the external ion source (positive ion mode), ionized upon irradiance with a 337 nm wavelength laser pulse, and transferred via rf-only quadrupole rods into a 6.5 Tesla magnetic field
- 25 where they were trapped collisionally. After a 15 second delay, ions were excited by a broadband chirp pulse and detected using 256K data points, resulting in time domain signals of 5 s duration. Reported (neutral) masses are those of the most abundant isotope peak after subtracting the mass of the charge carrying proton (1.01 Da).

-158-

Results

The first scheme presented utilizes the PROBE reaction shown schematically in Figure 65. A 20-mer primer is designed to bind specifically to a region on the complementary template downstream of the mutation site; upon annealing to the template, which is labelled with biotin and immobilized to streptavidin coated magnetic beads, the PROBE primer is presented with a mixture of the three deoxynucleotide triphosphates (dNTPs), a di-dNTP (ddNTP), and a DNA polymerase (Figure 65). The primer is extended by a series of bases specific to the identity of the variable base in codon 634; for any reaction mixture (e.g., ddA + dT + dC + dG), three possible extension products representing the three alleles are possible (Figure 65).

For the negative control (Figure 66), the PROBE reaction with ddATP + dNTPs (N = T, C, G) causes a $M_r(\text{exp})$ shift of the primer from 6135 to 6726 Da ($\Delta m = +591$). The absence of a peak at 6432 rules out a C→A mutation (Figure 65); the mass of the single observed peak is more consistent with extension by C-ddA ($M_r(\text{calc})$ 6721, +0.07% error) than by T-ddA ($M_r(\text{calc})$ 6736, -0.15% error) than of A_3TC_2G expected for C→A mutant. Combining the ddA and ddT reaction data, it is clear that the negative control is as expected homozygous normal at codon 634.

The ddA reaction for patient 1 also results in a single peak ($M_r(\text{exp}) = 6731$) between expected values for wildtype and C→T mutation (Figure 65b). The ddT reaction, however, results in two clearly resolved peaks consistent with a heterozygote wildtype ($M_r(\text{exp})$ 8249, +0.04% mass error)/C→T mutant ($M_r(\text{exp})$ 6428 Da, +0.08% mass error). For patient 2, the pair of Figure 66c ddA products represent a heterozygote C→A ($M_r(\text{exp})$ 6431, -0.06% mass error)/normal ($M_r(\text{exp})$ 6719, -0.03% mass error) allele. The ddT reaction confirms this, with a

-159-

single peak measured at 8264 Da consistent with unresolved wildtype and C→A alleles. The value of duplicate experiments is seen by comparing Figures 66a and 66b; while for patient 1 the peak at 6726 from the ddA reaction represents only one species, similar peak from
5 patient 1 is actually a pair of unresolved peaks differing in mass by 15 Da.

An alternate scheme for point mutation detection is differentiation of alleles by direct measurement of diagnostic product masses. A 44-mer containing the RET634 site was generated by the PCR, and the 19-
10 mer sense primer removed by NH_4OH cleavage at a ribonucleotide at its 3' terminus.

Figure 67 shows a series of MALDI-FTMS spectra of synthetic analogs of short amplified products containing the RET634 mutant site. Figures 67a-c and 67d-f are homozygous and heterozygous genotypes,
15 respectively. An internal calibration was done using the most abundant isotope peak for the wildtype allele; application of this (external) calibration to the five other spectra resulted in better than 20 ppm mass accuracy for each. Differentiation by mass alone of the alleles is straightforward, even for heterozygote mixtures whose components
20 differ by 16.00 (Figure 67d), 2501 (Figure 67e), or 9.01 Da (Figure 65f). The value of high performance MS is clear when recognition of small DNA mass shifts is the basis for diagnosis of the presence or absence of a mutation. The recent reintroduction of delayed extraction (DE) techniques has improved the performance of MALDI-TOF with shorts
25 DNAs (Roskey, M.T. *et al.*, (1996) *Anal. Chem.* 68:941-946); a resolving power (RP) of $> 10^3$ has been reported for a mixed-base 50-mer, and a pair of 31-mers with a C or a T (Δm 15 Da) at a variable position resolved nearly to baseline. Thus DE-TOF-MS has demonstrated the RP required for separation of the individual components of heterozygotes.

-160-

Even with DE, however, the precision of DNA mass measurement with TOF is typically 0.1% (8 Da at 8 kDa) using external calibration, sufficiently high to result in incorrect diagnoses. Despite the possibility of space charge induced frequency shifts (Marshall, A.G. *et al.* (1991)

- 5 Anal. Chem. 63:215A-229A), MALDI-FTMS mass errors are rarely as high as 0.005% (0.4 Da at 8 KDa), making internal calibration unnecessary.

- The methods for DNA point mutation presented here are not only applicable to the analysis of single base mutations, but also to less
10 demanding detection of single or multiple base insertions or deletions, and quantification of tandem two, three, or four base repeats. The PROBE reaction yields products amenable to analysis by relatively low performance ESI or MALDI instrumentation; direct measurement of short amplified product masses is an even more direct means of mutation
15 detection, and will likely become more widespread with the increasing interest in high performance MS available with FTMS.

EXAMPLE 16

Immobilization of nucleic acids on solid supports via an acid-labile covalent bifunctional trityl linker

- 20 Aminolinked DNA was prepared and purified according to standard methods. A portion (10eq) was evaporated to dryness on a speedvac and suspended in anhydrous DMF/pyridine (9:1; 0.1 ml). To this was added the chlorotrityl chloride resin (1 eq, 1.05 μ mol/mg loading) and the mixture was shaken for 24 hours. The loading was checked by taking a
25 sample of the resin, detritylating this using 80% AcOH, and measuring the absorbance at 260nm. Loading was ca. 150pmol/mg resin.

In 80% acetic acid, the half-life of cleavage was found to be substantially less than 5 minutes--this compares with trityl ether-based approaches of half-lives of 105 and 39 minutes for *para* and *meta*

-161-

substituted bifunctional dimethoxytrityl linkers respectively. Preliminary results have also indicated that the hydroxy picolinic acid matrix alone is sufficient to cleave the DNA from the chlorotrityl resin.

EXAMPLE 17

5 **Immobilization of nucleic acids on solid supports via hydrophobic trityl linker**

The primer contained a 5'-dimethoxytrityl group attached using routine trityl-on DNA synthesis.

10 CI8 beads from an oligo purification cartridge (0.2 mg) placed in a filter tip was washed with acetonitrile, then the solution of DNA (50 ng in 25 μ l) was flushed through. This was then washed with 5% acetonitrile in ammonium citrate buffer (70 mM, 250 μ l). To remove the DNA from the CI8, the beads were washed with 40% acetonitrile in water (10 μ l) and concentrated to ca 2 μ l on the Speedvac. The sample
15 was then submitted to MALDI.

The results showed that acetonitrile/water at levels of ca. > 30% are enough to dissociate the hydrophobic interaction. Since the matrix used in MALDI contains 50% acetonitrile, the DNA can be released from the support and successfully detected using MALDI-TOF MS (with the
20 trityl group removed during the MALDI process).

Figure 69 is a schematic representation of nucleic acid immobilization via hydrophobic trityl linkers.

EXAMPLE 18

25 **Immobilization of nucleic acids on solid supports via Streptavidin-Iminobiotin**

Experimental Procedure

2-iminobiotin N-hydroxy-succinimid ester (Sigma) was conjugated to the oligonucleotides with a 3'- or 5'-amino linker following the conditions suggested by the manufacturer. The completion of the

-162-

reaction was confirmed by MALDI-TOF MS analysis and the product was purified by reverse phase HPLC.

For each reaction, 0.1 mg of streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin from Dynal) were incubated with 80 pmol of the corresponding oligo in the presence of 1M NaCl and 50 mM ammonium carbonate (pH 9.5) at room temperature for one hour. The beads bound with oligonucleotides were washed twice with 50 mM ammonium carbonate (pH 9.5). Then the beads were incubated in 2 μ l of 3-HPA matrix at room temperature for 2 min. An aliquot of 0.5 μ l of supernatant was applied to MALDI-TOF. For biotin displacement experiment, 1.6. mol of free biotin (80-fold excess to the bound oligo) in 1 μ l of 50 mM ammonium citrate was added to the beads. After a 5 min. incubation at room temperature, 1 μ l of 3-HPA matrix was added and 0.5 μ l of supernatant was applied to MALDI-TOF MS. To maximize the recovery of the bound iminobiotin oligo, the beads from the above treatment were again incubated with a 2 μ l of 3-HPA matrix and 0.5 μ l of supernatant was applied to MALDI-TOF MS. The matrix alone and free biotin treatment quantitatively released iminobiotin oligo off the streptavidin beads as shown in Figures 70 and 71.

20

EXAMPLE 19**Mutation Analysis Using Loop Primer Oligo Base Extension****MATERIALS AND METHODS**

Genomic DNA. Genomic DNA was obtained from healthy individuals and patients suffering from sickle cell anemia. The wildtype and mutated sequences have been evaluated conventionally by standard Sanger sequencing.

25

PCR-Amplification. PCR amplifications of a part of the β -globin was established and optimized to use the reaction product without a further purification step for capturing with streptavidin coated bead. The

-163-

- target amplification for LOOP-PROBE reactions were performed with the loop-cod5 d(GAG TCA GGT GCG CCA TGC CTC AAA CAG ACA CCA TGG CGC, SEQ ID No. 58) as forward primer and β -11-bio d(TCT CTG TCT CCA CAT GCC CAG, SEQ ID. No. 59) as biotinylated reverse
- 5 primer. The underlined nucleotide in the loop-cod5 primer is mutated to introduce an invariant CfoI restriction site into the amplicon and the nucleotides in italics are complementary to a part of the amplified product. The total PCR volume was 50 μ l including 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, Cat# 1596594), 1.5
- 10 mM MgCl₂, 0.2 mM dNTPs (Boehringer-Mannheim, Ca# 1277049), and 10 pmol of each primer. A specific fragment of the β -globin gene was amplified using the following cycling condition: 5 min 94°C followed by 40 cycles of : 30 sec @ 94°C, 30 sec @ 56°C, 30 sec @ 72°C, and a final extension of 2 min at 72°C.
- 15 *Capturing and denaturation of biotinylated templates.* 10 μ l paramagnetic beads coated with streptavidin (10mg/ml; Dynal, Dynabeads M-280 streptavidin Cat# 112.06) and treated with 5x binding solution (5 M NH₄Cl, 0.3M NH₄OH) were added to 40 μ l PCR volume (10 μ l of the amplified product was saved for check electrophoresis).
- 20 After incubation for 30 min at 37°C the supernatant was discarded. The captured templates were denatured with 50 μ l 100 mM NaOH for 5 min at ambient temperature, then washed once with 50 μ l 50 mM NH₄OH and three times with 100 μ l 10mM Tris.Cl, pH 8.0. The single stranded DNA served as templates for PROBE reactions.
- 25 *Primer oligo base extension (PROBE) reaction.* The PROBE reactions were performed using Sequenase 2.0 (USB Cat# E70775Z including buffer) as enzyme and dNTPs and ddNTPs supplied by Boehringer-Mannheim (Cat# 1277049 and 1008382). The ratio between dNTPs (dCTP, dGTP, dTTP) and ddATP was 1:1 and the total used

-164-

concentration was 50 μM of each nucleotide. After addition of 5 μl 1-fold Sequenase-buffer the beads were incubated for 5 min at 65°C and for 10 min at 37°C. During this time the partially self complementary primer annealed with the target site. The enzymatic reaction started after

- 5 addition of 0.5 μl 100 mM dithiothreitol (DTT), 3.5 μl dNTP/ddNTP solution, and 0.5 μl Sequenase (0.8 U) and incubated at 37°C for 10 min. Hereafter, the beads were washed once in 1-fold TE buffer (10 mM Tris, 1mM EDTA, pH 8.0).

- Cfol restriction digest.* The restriction enzyme digest was
10 performed in a total volume of 5 μl using 10 U Cfol in 1-fold buffer L purchased from Boehringer-Mannheim. The incubation time was 20 min at 37°C.

Conditioning of the diagnostic products for mass spectrometric analysis

- 15 After the restriction digest, the supernatant was precipitated in 45 μl H_2O , 10 μl 3M NH_4^+ acetate (pH 6.5), 0.5 μl glycogen (10 mg/ml in water, Sigma, Cat# G1765), and 110 μl absolute ethanol for 1 hour at room temperature. After centrifugation at 13,000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 2 μl 18 Mohm/cm H_2O .
20 The beads were washed in 100 μl 0.7 M NH_4 citrate followed by 100 μl 0.05 M NH_4 citrate. The diagnostic products were obtained by heating the beads in 2 μl 50 mM NH_4OH at 80°C for 2 min.

Sample preparation and analysis on MALDI-TOF mass spectrometry.

- 25 Same preparation was performed by mixing 0.6 μl of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$) with 0.3 μl of either resuspended DNA/glycogen pellet or supernatant after heating the beads in 50 mM NH_4OH on a sample target and allowed to air dry. The sample target

-165-

was automatically introduced in to the source region of an unmodified Perspective Voyager MALDI-TOF operated in delayed extraction linear mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical molecular mass ($M_r(\text{calc})$) were calculated from atomic compositions; reported experimental ($M_r(\text{exp})$) values are those of the singly-protonated form.

RESULTS

The LOOP-PROBE has been applied to the detection of the most common mutation of codon 6 of the human β -globin gene leading to sickle cell anemia. The single steps of the method are schematically presented in figure 72. For the analysis of codon 6, a part of the β -globin gene was amplified by PCR using the biotinylated reverse primer $\beta 11\text{bio}$ and the primer loop-cod5 which is modified to introduce a *CfoI* recognition site (fig. 72a). The amplified product is 192 bp in length.

After PCR the amplification product was bound to streptavidin coated paramagnetic particles as described above. The antisense strand was isolated by denaturation of the double stranded amplified product (Fig. 72b). The intra-molecule annealing of the complementary 3' end was accomplished by a short heat denaturation step and incubation at 37°C.

The 3' end of the antisense strand is now partially double stranded (Fig. 72c). For analyzing the DNA downstream of the self annealed 3'-end of the antisense strand, the primer oligo base extension (PROBE) has been performed using ddATP, dCTP, dGTP, dTTP (Fig. 72d). This generates different products in length specific for the genotype of the analyzed individual. Before the determination of the length of these diagnostic products, the DNA was incubated with the *CfoI* restriction endonuclease that cuts 5' of the extended product. This step frees the stem loop from the template DNA whereas the extended product still keeps attached to

-166-

the template. The extended products are then denatured by heating from the template stand and analyzed by MALDI-TOF mass spectrometry.

Since the MALDI-TOF analyses were performed with a non-calibrated instrument, the mass deviation between observed and
5 expected values was approximately 0.6% higher than theoretically calculated. Nevertheless, the results obtained were conclusive and reproducible within repeated experiments. In all analyzed supernatants after the restriction digest the stem loop could be detected. Independent of the genotype, the stem loop has had in all analyses molecular masses
10 about 8150 Da (expected 8111 Da). An example is shown in Figure 73a. The second peak in this figure with a mass of 4076 Da is a doubly charged ion of the stem loop. Figure 73b to 73d show the analyses of different genotypes as indicated in the respective inserts. HbA is the wildtype genotype and HbC and HbS are two different mutations in
15 codon 6 of the β -globin gene which cause sickle cell disease. In the wildtype situation a single peak with a molecular mass of 4247 Da and another with 6696 Da are detected (Fig. 73b). The latter corresponds to the biotinylated PCR primer (β -11-bio) unused in the PCR reaction which also has been removed in some experiments. The former corresponds to
20 the diagnostic product for HbA. The analyses of the two individual DNA molecules with HbS trait as well as compound heterozygosity (HbS/HbC) for the sickle cell disorder lead also to unambiguous expected results (Fig. 73c and 73d).

In conclusion, the LOOP-PROBE is a powerful means for detection
25 of mutations especially predominant disease causing mutations or common polymorphisms. The technique eliminates one specific reagent for mutation detection and, therefore, simplifies the process and makes it more amenable to automation. The specific extended product that is analyzed is cleaved off from the primer and is therefore shorter compared

-167-

to the conventional method. In addition, the annealing efficiency is higher compared to annealing of an added primer and should therefore generate more product. The process is compatible with multiplexing and various detection schemes (e.g., single base extension, oligo base extension and sequencing). For example, the extension of the loop-primer can be used for generation of short diagnostic sequencing ladders within highly polymorphic regions to perform, for example, HLA typing or resistance as well as species typing (e.g., *Mycobacterium tuberculosis*)).

EXAMPLE 20

10 T7-RNA Polymerase Dependent Amplification of CKR-5 and Detection by MALDI-MS

MATERIALS AND METHODS

Genomic DNA. Human genomic DNA was obtained from healthy individuals.

- 15 *PCR-Amplification and Purification.* PCR amplification of a part of the CKR-5 gene was accomplished using ckrT7f as sense primer d(ACC TAG CGT TCA GTT CGA CTG AGA TAA TAC GAC TCA CTA TAG CAG CTC TCA TTT TCC ATA C (SEQ ID NO. 60). The underlined sequence corresponds to the sequence homologous to CKR-5, the bolded sequence corresponds to the T7-RNA polymerase promoter sequence and the italic sequence was chosen randomly. ckr5r was used as antisense primer d(AAC TAA GCC ATG TGC ACA ACA (SEQ ID NO. 61). Purification of the amplified product and removal of unincorporated nucleotides was carried out using the QIAquick purification kit (Qiagen, cat# 28104). In
- 20 the final PCR volume of 50 μ l were 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, cat# 1596594), 1.5 mM MgCl₂ 0.2 mM dNTPs (Boehringer-Mannheim, cat# 1277049), and 10 pmol of each primer. The specific fragment of the CKR-5 gene was amplified using the
- 25 following cycling conditions: 5 min @ 94°C followed by 40 cycles of 45

-168-

sec @ 94°C, 45 sec 52°C, 5 sec @ 72°C, and a final extension of 5 min at 72°C.

T7-RNA Polymerase conditions. One third of the purified DNA (about 60ng) was used in the T7-RNA polymerase reaction. (Boehringer-Mannheim, cat# 881 767). The reaction was carried out for 2h at 37°C according to the manufacturer's conditions using the included buffer. The final reaction volume was 20 μ l. 0.7 μ l RNasin (33 U/ μ l) had been added. After the extension reaction, the enzyme was inactivated by incubation for 5 min at 65°C.

10 *DNA digestion and conditioning of the diagnostic products for mass spec analysis.*

The template DNA was digested by adding RNase-free DNase I (Boehringer-Mannheim, cat# 776 758) to the inactivated T7 mixture and incubation for 20 min at room temperature. Precipitation was carried out by adding 1 μ l glycogen (10 mg/ml, Sigma, cat# G1765), 1/10 volume 3M NH₂acetate (pH 6.5), and 3 volume absolute ethanol and incubation for 1 hour at room temperature. After centrifugation at 13,000 g for 10 min, the pellet was washed in 70% ethanol and resuspended in 3 μ l 18 Mohm/cm H₂O. 1 μ l was analyzed on an agarose gel.

20 *Sample preparation and analysis on MALDI-TOF mass spectrometry*

Sample preparation was performed by mixing 0.6 μ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 H₂O:CH₃CN) with 0.3 μ l of resuspended DNA/glycogen on a sample target and allowed to air dry. The sample target was introduced into the source region of an unmodified Finnigan VISION2000 MALDI-TOF operated in relectron mode with 5kV. The theoretical molecular mass was calculated from atomic composition; reported experimental values are those of singly-protonated form.

30 **RESULTS**

-169-

The chemokine receptor CKR-5 has been identified as a major coreceptor in HIV-1 (see e.g., WO 96/39437 to Human Genome Sciences; Cohen, J. *et al.* Science 275:1261). A mutant allele that is characterized by a 32 bp deletion is found in 16% of the HIV-1 seronegative population whereas the frequency of this allele is 35% lower in the HIV-1 seropositive population. It is assumed that individuals homozygous for this allele are resistant to HIV-1. The T7-RNA polymerase dependent amplification was applied to identify this specific region of the chemokine receptor CKR-5 (Figure 74). Human genomic DNA was amplified using conventional PCR. The sense primer has been modified so that it contains a random sequence of 24 bases that facilitate polymerase binding and the T7-RNA polymerase promoter sequence (Figure 75). The putative start of transcription is at the first base 5' of the promoter sequence. ckr5r was used as an antisense primer. PCR conditions are outlined above. The amplified product derived from wildtype alleles is 75 bp in length. Primer and nucleotides were separated from the amplification product using the Qiagen QIAquick purification kit. One third of the purified product was applied to in vitro transcription with T7-RNA polymerase. To circumvent interference of the template DNA, it was digested by adding RNase-free DNase I. RNA was precipitated and this step also leaves the degraded DNA in the supernatant. Part of the redissolved RNA was analyzed on an agarose gel and the rest of the sample was prepared for MALDI-TOF analysis. The expected calculated mass of the product is 24560 Da. A dominant peak, that corresponds to an approximate mass of 25378.5 Da can be observed. Since the peak is very broad, an accurate determination of molecular mass was not possible. The peak does not correspond to residual DNA template. First, the template DNA is digested, and second,

-170-

the DNA strands would have a mass of 23036.0 and 23174 Da, respectively.

This example shows that T7 RNA polymerase can effectively amplify target DNA. The generated RNA can be detected by Mass
5 spectrometry. In conjunction with modified (e.g., 3'-deoxy)ribonucleotides that are specifically incorporated by a RNA polymerase but not extended any further, this method can be applied to determine the sequence of a template DNA.

EXAMPLE 21

10 MALDI Mass Spectrometry of RNA Endonuclease Digests

MATERIALS

Synthetic RNA (Sample A:5'-UCCGGUCUGAUGAGUCCGUGAGGAC-3' (SEQ ID 62); sample B:5'-GUCACUACAGGUGAGCUCCA-3' (SEQ ID NO 63); sample C:5'-
15 CCAUGCGAGAGUAAGUAGUA-3' (SEQ ID NO. 64)) samples were obtained from DNA technology (Aarhus, Denmark) and purified on a denaturing polyacrylamide gel (Shaler, T. A. *et al.* (1996) Anal. Chem. 63:5766-579). Rnases T₁ (Eurogentec), U₂ (Calbiochem), A (Boehringer-Mannheim) and PhyM (Pharmacia) were used without additional
20 purification. Streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Dynal) were supplied as a suspension of $6-7 \times 10^8$ bead/ml (10 mg/ml) dissolved in phosphate-buffered saline (PBS) containing 0.1% BSA and 0.02% NaN₃. 3-Hydroxypicolinic acid (3-HPA) (Aldrich) was purified by a separate desalting step before use as described in more
25 detail elsewhere (Little, D. P. *et al.* (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 2318-2322).

-171-

METHODS

In vitro transcription reaction. The 5'-biotinylated 49 nt in vitro transcript (SEQ ID No. 65):

AGGCCUGCGGCAAGACGGAAAGACCAUGGUCCCUNAUCUGCCGCAGGAUC

- 5 was produced by transcription of the plasmid pUTMS2 (linearized with the restriction enzyme BamHI) with T7 RNA polymerase (Promega). For the transcription reaction 3 μ g template DNA and 50u T7 RNA polymerase were used in a 50 μ l volume of 1u/ μ l RNA guard (Rnax inhibitor, Pharmacia), 0.5 mM NTP's 1.0 mM 5'-biotin-ApG dinucleotide, 10 40 mM Tris-HCl (pH 8.0), 6 mM MgCl₂ 2 mM spermidine and 10 mM DTT. Incubation was performed at 37°C for 1 hour, then another aliquot of 50 units T7 RNA polymerase was added and incubation was continued for another hour. The mixture was adjusted to 2M NH₄ acetate and the RNA was precipitated by addition of one volume of ethanol and one 15 volume of isopropanol. The precipitated RNA was collected by centrifugation at 20,000 X g for 90 min at 4°C, the pellet was washed with 70% ethanol, dried and redissolved at 8 M urea. Further purification was achieved by electrophoresis through a denaturing polyacrylamide gel as described elsewhere (Shaler, T. A. *et al.* (1996) 20 Anal. Chem. 68:576-579). The ration of 5'-biotinylated to non-biotinylated transcripts was about 3:1.

- Ribonuclease assay.* For partial digestion with selected RNases different enzyme concentrations and assay conditions were employed as summarized in table VII. The solvents for each enzyme were selected 25 following the suppliers' instructions. The concentrations of the synthetic RNA samples and the *in vitro* transcript were adjusted to 5-10 x 10⁻⁶M.

TABLE VII
Overview and Assay Conditions of the RNases

Rnase	Source	Concentration [units Rnase/ μ gRNA]	Conditions	Incubation Time (max. number of fragments)	References
T ₁	<i>Aspergillus oryzae</i>	0.2	20 mM Tris-HCl, pH 5.7, 37°C	5 min.	Donis-Keller, H. et al., (1977) Nuc. Acids Res. 4:2527- 2537
U ₂	<i>Ustilago Sphaerogena</i>	0.01	20 mM DAC, pH 5.0, 37°C	15 min	Donis-Keller, H. et al., (1977) Nuc. Acids Res. 4:2527- 2537
PhyM	<i>Physarum polycephalum</i>	20	20 mM DAC, pH 5.0, 50°C	15 min	Donis-Keller, H. et al., (1980) Nucl. Acids Res. 8:3133- 3142
A	bovine pancreas	4×10^{-9}	10 mM Tris-HCl, pH 7.5, 37°C	30 min	Breslow, R. and R. Xu. (1993) Proc. Natl. Acad. Sci. USA 90:1201-1207
CL ₃	chicken liver	0.01	10 mM Tris-HCl, pH 6.5, 37°C	30 min	Boguski, et al., (1980) J. Biol. Chem. 255:2160-2163
cusativin	<i>cucumis sativus</i> L.	0.05 ng	10 mM Tris-HCl, pH 6.5, 37°C	30 min	Rojo, M.A. et al. (1994) Planta 194:328-338

-173-

The reaction was stopped at selected times by mixing 0.6 μ l aliquots of the assay with 1.5 μ l of 3 HPA-solution. The solvent was subsequently evaporated in a stream of cold air for the MALDI-MS analysis.

- 5 Limited alkaline hydrolysis was performed by mixing equal volumes (2.0 μ l) of 25% ammonium hydroxide and RNA sample ($5-10 \times 10^{-6}$ M) at 60°C. 1 μ l aliquots were taken out at selected times and dried in a stream of cold air. For these samples it turned out to be important to first dry the digests in a stream of cold air, before 1.5 μ l of the matrix solution and 0.7 μ l of NH_4^+ loaded cation exchanged polymer beads were added.

- 10 The reaction was stopped at selected times by mixing 0.6 μ l aliquots of the assay with 1.5 μ l of 3HPA-solution. The solvent was subsequently evaporated in a stream of cold air for the MALDI-MS analysis.

- 15 Limited alkaline hydrolysis was performed by mixing equal volumes (2.0 μ l) of 25% ammonium hydroxide and RNA sample ($5-10 \times 10^{-6}$ M) at 60°C. 1 μ l aliquots were taken out at selected times and dried in a stream of cold air. For these samples it turned out to be important to first dry the digests in a stream of cold air, before 1.5 μ l of the matrix solution and 0.7 μ l if a suspension of NH_4^+ loaded cation exchange polymer beads were added.

- 20 *Separation of 5'-biotinylated fragments.* Streptavidin-coated magnetic beads were utilized to separate 5'-biotinylated fragments of the *in vitro* transcript after partial RNase degradation. The biotin moiety in

-174-

- this sample was introduced during the transcription reaction initiated by the 5'-biotin-pApG-dinucleotide. Prior to use, the beads were washed twice with 2 x binding & washing (b&w) buffer (20 mM Tris-HCl, 2 mM EDTA, 2 M NaCl pH 8.2) and resuspended at 10 mg/ml in 2 x b&w
- 5 buffer. Circa 25 pmol of the RNA *in vitro* transcript were digested by RNase U2 using the protocol described above. The digestion was stopped by adding 3 μ l of 95% formamide containing 10 mM trans- 1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid (CDTA) at 90°C for 5 min, followed by cooling on ice. Subsequently, capture of the
- 10 biotinylated fragments was achieved by incubation of 6 μ l of the digest with 6 μ l of the bead suspension and 3 μ l of b&w buffer at room temperature for 15 min. Given the binding capacity of the beads of 200 pmol of biotinylated oligonucleotide per mg of beads, as specified by the manufacturer, the almost 2-times excess of oligonucleotide was used to
- 15 assure a full loading of the beads. The supernatant was removed, and the beads were washed twice with 6 μ l of H₂O. The CDTA and 95% formamide at 90°C for 5 min. After evaporation of the solvent and the formamide the ≤ 2.5 pmol of fragments were resuspended in 2 μ l H₂O and analyzed by MALDI-MS as described above.
- 20 *Sample preparation for MALDI-MS.* 3-Hydroxypicolinic acid (3-HPA) was dissolved in ultra pure water to a concentration of ca. 300 mM. Metal cations were exchanged against NH₄⁺ as described in detail previously. (Little, D. P. *et al.* (1995)Proc. Natl. Acad. Sci. U.S.A. 92: 2318-2322). Aliquots of 0.6 μ l of the analyte solution were mixed with
- 25 1.5 μ l 3-HPA on a flat inert metal substrate. Remaining alkali cations,

-175-

present in the sample solution as well as on the substrate surface, were removed by the addition of 0.7 μ l of the solution of NH_4^+ - loaded cation exchange polymer beads. During solvent evaporation, the beads accumulated in the center of the preparation, were not used for the
5 analysis, and were easily removed with a pipette tip.

Instrument. A prototype of the Vision 2000 (ThermBioanalysis, Hemel, Hempstead, UK) reflectron time of flight mass spectrometer was used for the mass spectrometry. Ions were generated by irradiation with a frequency-tripled ND:YAG laser (355 nm, 5 ns; Spektrum GmbH,
10 Berlin, Germany) and accelerated to 10 ke V. Delayed ion extraction was used for the acquisition of the spectra shown, as it was found to substantially enhance the signal to noise ratio and/or signal intensity. The equivalent flight path length of the system is 1.7 m, the base pressure is 10^{-4} Pa. Ions were detected with a discrete dynode
15 secondary-electron multiplier (R2362, Hamamatsu Photonics), equipped with a conversion dynode for effective detection of high mass ions. The total impact energy of the ions on the conversion dynode was adjusted to values ranging from 16 to 25 keV, depending on the mass to be detected. The preamplified output signal of the SEM was digitized by a
20 LeCroy 9450 transient recorder (LeCroy, Chestnut Ridge, NY, USA) with a sampling rate of up to 400 MHz. For storage and further evaluation, the data were transferred to a personal computer equipped with custom-made software (ULISSES). All spectra shown were taken in the positive ion mode. Between 20 and 30 single shot spectra were averaged for
25 each of the spectra shown.

-176-

RESULTS

Specificity of Rnases. Combining base-specific RNA cleavage with MALDI-MS requires reaction conditions optimized to retain the activity and specificity of the selected enzymes on the one hand and complying with the boundary conditions for MALDI on the other. Incompatibility mainly results because the alkaline-ion buffers, commonly used in the described reaction, such as Na-phosphate, Na-citrate or Na-acetate as well as EDTA interfere with the MALDI sample preparation; presumably they disturb the matrix crystallization and/or analyte incorporation. Tris-HCl or ammonium salt buffers, in contrast, are MALDI compatible (Shaler, T. A. *et al.* (1996) Anal. Chem. **68**:576-579). Moreover, alkaline salts in the sample lead to the formation of a heterogenous mixture of multiple salts of the analyte, a problem increasing with increasing number of phosphate groups. Such mixtures result in loss of mass resolution and accuracy as well as signal-to-noise ratio (Little, D. P. *et al.* (1995) Proc. Natl. Acad. Sci. U.S.A. **92**:2318-2322; Nordhoff, E., Cramer, R. Karas, M., Hillenkamp, F., Kirpekar, F., Kristiansen, K. and Roepstorff, P. (1993) Nucleic Acids Res., **21**, 3347-3357). Therefore, RNase digestions were carried out under somewhat modified conditions compared to the ones described in the literature. They are summarized above in table VII. For Rnase T₁, A, CL₃ ad Cusativin, Tris-HCl (pH 6-7.5) was used as buffer. 20 mM DAC provides the pH of 5, recommended for maximum activity of RNases U₂ and PhyM. The concentration of 10-20 mM of these compounds were found to not interfere significantly with the MALDI analysis. To examine the specificity of the selected ribonucleases under these conditions, three synthetic 20-25mer RNA molecules with different nucleotide sequences were digested.

-177-

The MALDI-MS spectra of Figure 77 shows five different cleavage patterns (A-E) of a 25 nt RNA obtained after partial digestion with RNases T₁, U₂, PhyM, A, and alkaline hydrolysis. These spectra were taken from aliquots which were removed from the assay after empirically
5 determined incubation times, chosen to get an optimum coverage of the sequence. As the resulting samples were not fractionated prior to mass spectrometric analysis, they contain all fragments generated at that time by the respective RNases. In practice, uniformity of the cleavages, can be affected by a preferential attack on the specific phosphodiester bonds
10 (Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) *Nucleic Acids Res.*, **4**, 1957-1978; Donis-Keller, H. (1980) *Nucleic Acids Res.*, **8** 3133-3142). The majority of the expected fragments are indeed observed in the spectra. It is also worth noting that for the reaction protocols as used, correct assignment of all fragment masses is only possible, if a 2',
15 3'-cyclic phosphate group is assumed. It is well known that such cyclic phosphates are intermediates in the cleavage reaction and get hydrolyzed in a second, independent and slower reaction step involving the enzyme (Richards, F. M., and Wycoff, H. W. in *The Enzymes* Vol. 4, 3rd Ed., (ed. Boyer, P.D.) 746-806 (1971, Academic Press, New York); Heinemann, U
20 and W. Saenger (1985) *Pure Appl. Chem.* **57**, 417-422; Ikehara, M. et al., (1987) *Pure Appl. Chem.* 59-965-968) Vreslow, R. and Xu, R. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1201-1207). In a few cases different fragments have equal mass or differ by as little as 1 Dalton., In these cases, mass peaks cannot unambiguously be assigned to one or the
25 other fragments. Digestion of two additional different 20 nt RNA samples was, therefore, performed (Hahner, S., Kirpekar, F., Nordoff, E., Kristiansen, K., Roepstorff, P. and Hillenkamp, F. (1996) *Proceedings of the 44th ASMS Conference on Mass Spectrometry*, Portland, Oregon) in order to sort out these ambiguities. For all samples tested, the selected

-178-

ribonucleases appear to cleave exclusively at the specified nucleotides leading to fragments arising from single as well as multiple cleavages.

In Figure 77, peaks, indicating fragments containing the original 5'-terminus, are marked by arrows. All non marked peaks can be assigned to internal sequences or those with retained 3'-terminus. For a complete sequence all possible fragments bearing exclusively either the 5'- or the 3'-terminus of the original RNA would suffice. In practice, the 5'-fragments are better suited for this purpose, because the spectra obtained after incubation of all three synthetic RNA samples contain the nearly complete set of originals of 5'-ions for all different RNases (Hahner, S., Kirpekar, F., Nordoff, E., Kristiansen, K., Roepstorff and Hillenkamp, F. (1996) *Proceedings of the 44th ASMS Conference on Mass Spectrometry*, Portland, Oregon). Internal fragments are somewhat less abundant and fragments containing the original 3'-terminus appear suppressed in the spectra. In agreement with observations reported in the literature (Gupta, R. C. and Randerath, K. (1977) *Nucleic Acids Res.*, **4**, 1957-1978), cleavages close to the 3'-terminus were partially suppressed in partial digests of the RNA 25 mer by RNase T₁ and U₂ (even if they are internal or contain the original 5'-terminus). Fragments from such cleavages appear as weak and poorly resolved signals in the mass spectra.

For larger RNA molecules secondary structure is known to influence the uniformity of the enzymatic cleavages (Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) *Nucleic Acids Res.* **5**, 3133-3142). This can, in principle be, overcome by altered reaction conditions. In assay solutions containing 5-7 M urea, the activity of RNases such as T₂, U₂, A, C₁, and PhyM is known to be retained (Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) *Nucleic Acids Res.*, **4**, 2527-2537; Boguski, M. S., Hieter, P.A., and Levy, C. C. (1980) *J. Biol. Chem.*, **255**,

-179-

2160-2163; Donis-Keller, H. (1980) *Nucleic Acids Res.*, **8**, 3133-3142, while RNA is sufficiently denatured. UV-MALDI-analysis with 3-HPA as matrix is not possible under such high concentrations of urea in the sample. Up to a concentration of 2 M urea in the reaction buffer, MALDI
5 analysis of the samples was still possible although significant changes in matrix crystallization were observed. Spectra of the RNA 20 mer (sample B), digested in the presence of 2 M urea still resembled those obtained under conditions listed in Table VII.

Digestion by RNases which exclusively recognize one nucleobase
10 is desirable to reduce the complexity of the fragment patterns and thereby facilitate the mapping of the respective nucleobase. RNases CL₃ and cursavitin are enzymes reported to cleave at cytidylic acid residues. Upon limited RNase CL₃ and cursativin digestion of the RNA-20mer (sample B) under non-denaturing conditions, fragments corresponding to
15 cleavages at cytidylic residues were indeed observed (Figure 78). Similar to the data reported so far (Boguski, M. S., Hieter, P. A. and Levy, C. C. (1980) *J. Biol. Chem.*, **255**, 2160-2163; Rojo, M. A., Arias, F. J., Iglesias, R., Ferreras, J. M., Munoz, R., Escarmis, C., Soriano, F., Llopez-Fando, Mendez, E., and Girbes, T. (1994) *Planta*, **194**, 328-338). The
20 degradation pattern in Figure 78, however, reveals that not every cytidine residue is recognized, especially for neighboring C residues. RNase CL₃ is also reported to be susceptible to the influence of secondary structure (Boguski, M. S., Heiter, P. A., and Levy, C. C. (1980) *J Biol. Chem.*, **255**, 2160-2163), but for RNA of the size
25 employed in this study, such an influence should be negligible. Therefore, unrecognized cleavage sites in this case can be attributed to a lack of specificity of this enzyme. To confirm these data, a further RNase CL₃-digestion was performed with the RNA 20mer (sample C). As a result of the sequence of this analyte, all three linkages containing

-180-

- cytidylic acid were readily hydrolyzed, but additional cleavages at uridylic acid residues were detected as well. Since altered reaction conditions such as increased temperature (90°C), various enzyme to substrate ratios, and addition of 2M urea did not result in a digestion of the
- 5 expected specificity, application of this enzyme to sequencing was not pursued further. Introduction of a new cytidine-specific ribonuclease, cusativin, isolated from dry seeds of *Cucumis sativus* L. looked promising for RNA sequencing (Rojo, M. A., Arias, F. J., Iglesias, R., Ferreras, J. M., Munoz, R., Escarmis, C., Soriano, F., Llopez-Fando, J., Mendez, E.
- 10 and Girbes, T. (1994) *Planta*, **194**, 328-338). As shown in Figure 78, not every cytidine residue was hydrolyzed and additional cleavages occurred at uridylic acid residues for the recommended concentration of the enzyme. RNases CL₃ and cusativin will, therefore not yield the desired sequence information for mapping of cytidine residues and their
- 15 use was not further pursued. The distinction of pyrimidine residues can be achieved, however, by use of RNases with multiple specificities, such as *Physarum polycephalum* RNase (cleaves ApN, UpN) and pancreatic RNase A (cleaves UpN, CpN) (see Figure 77). All 5'-terminus fragments, generated by the monospecific RNase U₂ and apparent in the spectrum of
- 20 Figure 77C were also evident in the spectrum of the RNase PhyM digest (Figure 77D). Five of the six uridilic cleavage sites could, this way, be uniquely identified by this indirect method. In a next step, the knowledge of the uridine cleavage sites was used to identify sites of cleavage of cytidilic acid residues in the spectrum recorded after
- 25 incubation with RNase A (Figure 77E), again using exclusively ions containing the original 5'-terminus. Two of the four expected cleavage sites were identified this way. A few imitations are apparent from these spectra, if only the fragments containing the original 5'-terminus are used for the sequence determination. The first two nucleotides usually escape

-181-

the analysis, because their signals get lost in the low mass matrix background. Because of this, the corresponding fragments are missing in the spectra of the U- and C-specific cleavages. Large fragments with cleavage sites close to the 3'-terminus are often difficult to identify, particularly in digests with RNases T¹ and U₂, because of their low yield (vide supra) and the often strong nearby signal of the non-digested transcript. Accordingly the cleavages in position 22 and 23 do not show up in the spectrum of the G-specific RNase T, (Figure 77A) and the cleavage site 24 cannot be identified from the spectra of the U₂ and PhyM digests (Figures 77 C and D). Also site 16 and 17 with two neighboring cytidilic acids cannot be identified in the RNase A spectrum of Figure 77E. These observations demonstrate that a determination of exclusively the 5'-terminus fragments may not always suffice and the information contained in the internal fragments may be needed for a full sequence analysis.

Finally, limited alkaline hydrolysis provides a continuum of fragments (Figure 77B), which can be used to complete the sequence data. Again, the spectrum is dominated by ions of fragments containing the 5'-terminus, although the hydrolysis should be equal for all phosphodiester bonds. As was true for the enzymatic digests, correct mass assignments requires one to assume that all fragments have a 2', 3'-cyclic phosphate. The distribution of peaks, therefore, resembles that obtained after a 3'-exonuclease digest (Pieles, U., Zurcher, W., Schar, M. and Moser, H. E., (1993) *Nucleic Acids Res.*, **21**, 3191-3196; Nordhoff, E. et al. (1993) Book of Abstracts, 13th Internat. Mass Spectrom. Conf., Budapest p. 218; Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Lezius, A. Hahner, S., Karas, M. and Hillenkamp, F. (1994) *Nucleic Acid Res.*, **22**, 3866-3870). In principle, the alkaline hydrolysis alone could, therefore, be used for a complete sequencing. This is, however,

-182-

only possible for quite small oligoribonucleotides, because larger fragment ions, differing in mass by only a few mass units will not be resolved in the spectra and the mass of larger ions cannot be determined with the necessary accuracy of better than 1 Da, even if peaks are

5 partially or fully resolved. The interpretation of the spectra particularly from digests of unknown RNA samples is substantially simplified, if only the fragments containing the original 5'-terminus are separated out prior to the mass spectrometric analysis. A procedure for this approach is described in the following section.

- 10 *Separation of 5'-biotinylated fragments.* Streptavidin-coated magnetic beads (Dynal) were tested for the extraction of fragments containing the original 5'-terminus from the digests. Major features to be checked for this solid-phase approach are the selective immobilization and efficient elution of biotinylated species. In preliminary experiments,
- 15 a 5'-biotinylated DNA (19 nt) and streptavidin were incubated and MALDI analyzed after standard preparation. Despite the high affinity of the streptavidin-biotin interaction, the intact complex was not found in the MALDI spectra. Instead, signals of the monomeric subunit of streptavidin and the biotinylated DNA were detected. Whether the
- 20 complex dissociates in the acidic matrix solution (pKA 3) or during the MALDI desorption process, is not known. Surprisingly, if the streptavidin is immobilized on a solid surface such as magnetic beads, the same results are not observed. A mixture of two 5'-biotinylated DNA samples (19 nt and 27 nt) and two unlabeled DNA sequences (12 nt and 22 nt)
- 25 were incubated with the beads. The beads were extracted and carefully washed before incubation in the 3-HPA MALDI matrix. No analyte signals could be obtained from these samples. To test whether the biotinylated species had been bound to the beads altogether, elution from the extracted and washed beads was performed by heating at 90°C in

-183-

the presence of 95% formamide. This procedure is expected to denature the streptavidin, thereby breaking the streptavidin/biotin complex. Figure 79B shows the expected signals of the two biotinylated species, proving that release of the bound molecules in the MALDI process is the problem rather than the binding of the beads; Figure 79A shows a spectrum of the same sample after standard preparation, showing signals of all four analytes as a reference. Complete removal of the formamide after the elution and prior to the mass spectrometric analysis was found to be important, otherwise crystallization of the matrix is disturbed. Mass resolution and the signal-to-noise ration in spectrum 79B are comparable to those of the reference spectrum. These results testify to the specificity of the streptavidin-biotin interaction, since no or only minor signals of the non-biotinylated analyte were detected after incubation with the Dynal beads. Increased suppression of nonspecific binding was reported through an addition of the detergent Tween-20 to the binding buffer (Tong, X. and Smith, L. M. (1992) *Anal Chem.*, **64**, 2672-2677). Although this effect could be confirmed in this study, peak broadening affected the quality of the spectra due to remaining amounts of the detergent. The necessity of an elution step as a prerequisite for detection of the captured biotinylated species can be attributed to a stabilizing effect of the complex by the immobilization of the streptavidin to the magnetic beads.

For practical application of this solid phase method to sequencing a maximum efficiency of binding and elution of biotinylated species is of prime importance. Among a variety of conditions investigated so far, addition of salts such as EDTA gave best results in the case of DNA sequencing by providing ionic strength to the buffer (Tong, X. and Smith, L. M. (1992) *Anal Chem.*, **64**, 2672-2677). To examine such an effect on the solid-phase method, several salt additives were tested for the

-184-

binding and elution of the 5'-biotinylated RNA *in vitro* transcript (49 nt). The results are shown in Figure 80. Judging from the relative intensity, signal-to-noise ration, and resolution of the respective signals, a 95% formamide solution containing 10 mM CDTA (Figure 80D) is most efficient for the binding/elution. Since CDTA acts as a chelating agent for divalent cation, formation of proper secondary and tertiary structure of the RNA is prevented. An improved sensitivity and spectral resolution has been demonstrated under such conditions for the analysis of RNA samples by electrospray mass spectrometry (Limbach, P. A., Crain, P. F. and McCloskey, J.A. (1995) *J Am. Soc. Mass. Spectrom.*, 6, 27-39). The improvement in the MALDI analysis is actually not very significant compared to the spectrum obtained for the solution containing formamide alone (Figure 81b), but the reproducibility for spectra of good quality was substantially improved for the CDTA/formamide solution. Thus in addition to the improved binding/elution, this additive may also improve the incorporation of the analyte into the matrix crystals. Unfortunately, a striking signal broadening on the high mass side was observed in case of formamide solutions containing EDTA, CDTA or 25% ammonium hydroxide. Since this effect is most prominent in case of 25% ammonium hydroxide and this agent was also used for adjusting EDTA and CDTA to their optimum pH, a pronounced NH_3 adduct ion formation can be assumed.

The applicability of streptavidin-coated magnetic beads separation to RNA sequencing was demonstrated for the Rnase U_2 digest of the 5'-biotinylated RNA *in vitro* transcript (49 nt) (Figure 81). The entire fragment pattern obtained after incubation with Rnase U_2 is shown in spectrum 81A. Separation of the biotinylated fragments reduces the complexity of the spectrum (Figure 81B) since only 5'-terminal fragments are captured by the beads. The signals in the spectrum are broadened

-185-

and the increased number of signals in the low mass range indicate that even after stringent washing of the beads, some amounts of buffer and detergent used for the binding and elution remained. Further improvements of the method are, therefore, needed. Another possible
5 strategy for application of the magnetic beads is the immobilization of the target RNA prior to RNase digestion by an elution of the remaining fragments for further analysis. Cleavage of the RNA was impeded in this case, as evidenced by a prolonged reaction time for the digest under otherwise identical reaction conditions.

10

EXAMPLE 22**Parallel DNA Sequencing Mutation Analysis and Microsatellite Analysis Using Primers with Tags and Mass Spectrometric Detection**

This EXAMPLE describes specific capturing of DNA products generated in DNA analysis. The capturing is mediated by a specific tag
15 (5 to 8 nucleotides long) at the 5' end of the analysis product that binds to a complementary sequence. The capture sequence can be provided by a partially double stranded oligonucleotide bound to a solid support. Different DNA analysis (e.g., sequencing, mutation, diagnostic, microsatellite analysis) can be carried out in parallel, using, for example,
20 a conventional tube or microtiter plate (MTP). The products are then specifically captured and sorted out via the complementary identification sequence on the tag oligonucleotide. The capture oligonucleotide can be bound onto a solid support (e.g., silicon chip) by a chemical or biological bond. Identification of the sample is provided by the predefined position
25 of the capture oligonucleotide. Purification, conditioning and analysis by mass spectrometry are done on solid support. This method was applied for capturing specific primers that had a 6 base tag sequence.

MATERIALS AND METHODS*Genomic DNA.*

-186-

Genomic DNA was obtained from healthy individuals.

PCR Amplification

PCR amplifications of part of the β -globin gene were established using β 2 d(CATTTGCTTCTGACACAACT Seq. ID. No. 66) as forward
5 primer and β 11 d(TCTCTGTCTCCACATGCCAG Seq. ID. No. 67) as reverse primer. The total PCR volume was 50 μ l including 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, Cat# 159594), 1.5 mM $MgCl_2$, 0.2 mM dNTPs (Boehringer-Mannheim, Cat# 1277049), and 10 pmol of each primer. A specific fragment of the β -
10 globin gene was amplified using the following cycling conditions: 5 min @ 94°C followed by 40 cycles of 30 sec @ 94°C, 45 sec @ 53°C, 30 sec @ 72°C, and a final extension of 2 min @ 72°C. Purification of the amplified product and removal of unincorporated nucleotides was carried out using the QIAquick purification kit (Qiagen, Cat 28104). One fifth of
15 the purified product was used for the primer oligo base extension (PROBE) or sequencing reactions, respectively.

Primer oligo base extension (PROBE) and sequencing reactions

Detection of putative mutations in the human β -globin gene at codon 5 and 6 and at codon 30 and in the IVS-1 donor site, respectively,
20 was done in parallel (FIGURE 82A). β -TAG1 (GTCGTCCCATGGTGCACCTGACTC Seq. ID. No. 68) served as primer to analyze codon 5 and 6 and β -TAG2 (CGCTGTGGTGAGGCCCTGGGCA Seq. ID. No. 69) for the analyses of codon 30 and the IVS-1 donor site. The primer oligo base extension (PROBE) reaction was done by cycling,
25 using the following conditions: final reaction volume was 20 μ l, β -TAG1 primer (5 pmol), β -TAG2 primer (5 pmol), dCTP, dGTP, dTTP, (final concentration each 25 μ M), ddATP (final concentration 100 μ M) dNTPs and ddNTPs purchased from Boehringer-Mannheim, Cat# 1277049 and 1008382), 2 μ l of 10x ThermoSequence buffer and 2.5 U

-187-

ThermoSequenase (Amersham, CAT#E79000Y). The cycling program was as follows: 5 min @ 94°C, 30 sec @ 53°C, 30 sec @ 72°C and a final extension step for 8 min @ 72°C. Sequencing was performed under the same conditions except that the reaction volume was 25 μ l and the concentration of nucleotides was 250 μ M for ddNTP.

Capturing using TAG sequence and sample preparation

The capture oligonucleotides cap-tag1
d(GACGACGACTGCTACCTGACTCCA Seq ID No. 70) and cap-tag2
d(ACAGCGGACTGCTACCTGACTCCA Seq ID No. 71), respectively, were
annealed to equimolar amounts of uni-as d(TGGAGTCAGGTAGCAGTC
Seq ID No. 72) (FIGURE 82A). Each oligonucleotide had a concentration
of 10 pmol/ μ l in ddH₂O and incubated for 2 min @ 80°C and 5 min @
37°C. This solution was stored at -20°C and aliquots were taken. 10
pmol annealed capture oligonucleotides were bound to 10 μ l paramagnetic
beads coated with streptavidin (10 mg/ml; Dynal, Dynabeads M-280
streptavidin Cat# 112.06) by incubation for 30 min @ 37°C. Beads
were captured and the PROBE or sequencing reaction, respectively, was
added to the capture oligonucleotides. To facilitate binding of β -TAG1
and β -TAG2, respectively, the reaction was incubated for 5 min @ 25°C
and for 30 min @ 16°C. The beads were washed twice with ice cold
0.7 M NH₄ Citrate to wash away unspecific bound extension products
and primers. The bound products were dissolved by adding 1 μ l DDH₂O
and incubation for 2 min @ 65°C and cooling on ice. 0.3 μ l of the
sample were mixed with 0.3 μ l matrix solution (saturated 3-hydroxy-
picolinic acid, 10% molar ratio ammonium-citrate in acetonitrile/water
(50/50. v/v)) and allowed to air dry. The sample target was
automatically introduced into the source region of an unmodified
Perspective Voyager MALDI-TOF operated in delayed extraction linear
mode with 5 and 20 kV on the target and conversion dynode,

-188-

respectively. Theoretical average molecular mass ($M_r(\text{calc})$) were calculated from atomic compositions; reported experimental $M_r(M_r(\text{exp}))$ values are those of the singly-pronated form.

RESULTS

- 5 Specific capturing of a mixture of extension products by a short complementary sequence has been applied to isolate sequencing and primer oligo base extension (PROBE) products. This method was used for the detection of putative mutations in the human β -globin gene at codon 5 and 6 and at codon 30 and IVS-1 donor site, respectively
- 10 (FIGURE 82A). Genomic DNA has been amplified using the primers β 2 and β 11. The amplification product was purified and the nucleotides separated. One fifth of the purified product was used for analyses by primer oligo base extension. To analyze both sites in a single reaction, primers, β -TAG1 and β -TAG2, were used respectively. β -TAG1 binds
- 15 upstream of codons 5 and 6 and β -TAG2 upstream of codon 30 and the IVS-1 donor site. Extension of these primers was performed by cycling in the presence of ddATP and dCTP, dGTP and dTTP, leading to specific products, depending on the phenotype of the individual. The reactions were then mixed with the capture oligonucleotides. Capture
- 20 oligonucleotides include the biotinylated capture primer cap-tag1 and cap-tag2, respectively. They have 6 bases at the 5' end, that are complementary to the 5' end of β -TAG1 and β -TAG2, respectively. Therefore, they specifically capture these primers and the extended products. By annealing a universal oligonucleotide (uni-as) to the capture
- 25 oligonucleotide, the capture primer is transformed into a partially double stranded molecule where only the capture sequence stays single stranded (Figure 82). This molecule is then bound to streptavidin coated paramagnetic particles, to which the PROBE or sequencing reaction, respectively is added. The mixture was washed to bind only the

-189-

specifically annealed oligonucleotides. Captured oligonucleotides are dissolved and analyzed by mass spectrometry.

PROBE products of one individual (Fig. 83) show a small peak with a molecular mass of 7282.8 Da. This corresponds to the unextended β -TAG1 that has a calculated mass of 7287.8 Da. The peak at 8498.6 Da corresponds to a product, that has been extended by 4 bases. This corresponds to the wildtype situation. The calculated mass of this product is 8500.6 Da. There is no significant peak indicating a heterozygote situation. Furthermore only β -TAG1 and not β -TAG2 has been captured, indicating a high specificity of this method.

Analyses of what was bound to cap-tag2 (Figure 84) shows only one predominant peak with a molecular mass of 9331.5 Da. This corresponds to an extension of 8 nucleotides. It indicates a homozygous wildtype situation where the calculated mass of the expected product is 9355 Da. There is no significant amount of unextended primer and only β -TAG2 has been captured.

To prove that this approach is also suitable for capturing specific sequencing products, the same two primers β -TAG1 and β -TAG2, respectively, were used. The primers were mixed, used in one sequencing reaction and then sorted by applying the above explained method. Two different termination reactions using ddATP and ddCTP were performed with these primers (Figures 85 and 86, respectively). All observed peaks in the spectrograms correspond to the calculated masses in a wildtype situation.

As shown above, parallel analysis of different mutations (e.g., different PROBE primers) is now possible. Further, the described method is suitable for capturing specific sequencing products. Capturing can be used for separation of different sequencing primers out of one reaction tube/well, isolation of specific multiplex-amplified products, PROBE

-190-

products, etc. Conventional methods, like cycle sequencing, and conventional volumes can be used. A universal chip design permits the use of many different applications. Further, this method can be automated for high throughput.

5

EXAMPLE 23**Deletion Detection by Mass-Spectrometry**

Various formats can be employed for mass spectrometer detection of a deletion within a gene. For example, molecular mass of a double standard amplified product can be determined, or either or both of the
10 strands of a double stranded product can be isolated and the mass measured as described in previous examples.

Alternatively, as described herein, a specific enzymatic reaction can be performed and the mass of the corresponding product can be determined by mass spectrometry. The deletion size can be up to
15 several tenths of vases in length, still allowing the simultaneous detection of the wildtype and mutated allele. By simultaneous detection of the specific products, it is possible to identify in a single reaction whether the individual is homozygous or heterozygous for a specific allele or mutation.

20 MATERIALS AND METHODS*Genomic DNA*

Leukocyte genomic DNA was obtained from unrelated healthy individuals.

PCR amplification

25 PCR amplification of the target DNA was established and optimized to use the reaction products without a further purification step for capturing with streptavidin coated beads. The primers for target amplification and for PROBE reactions were as follows:

-191-

CKRA-F:d(CAG CTC TCA TTT TCC ATA C SEQ ID. NO. 73) and CKRA-R bio: d(AGC CCC AAG ATG ACT ATC SEQ ID. NO. 74). CKR-5 was amplified by the following program: 2 min @ 94°C, 45 seconds @ 52°C, 5 seconds @ 72°C, and a final extension of 5 minutes at 72°C.

- 5 The final volume was 50 μ l including 200 ng genomic DNA 1U Taq-polymerase (Boehringer-Mannheim, Cat # 1596594), 1.5 Mm $MgCl_2$, 0.2 Mm DNTPS (Boehringer-Mannheim, Cat # 1277049), 10 pmol of unmodified forward primers, and 8 pmol 5' biotinylated reverse primer.

Capturing and Denaturation of Biotinylated Templates

- 10 10 μ l paramagnetic beads coated with streptavidin (10 mg/ml; Dynal, Dynabeads M-280 streptavidin Cat # 112.06) in 5x binding solution (5M NH_4Cl , 0.3 M NH_4OH) were added to 45 μ l PCR reaction (5 μ l of PCR reaction were saved for electrophoresis). After binding by incubation for 30 min. at 37°C the supernatant was discarded. Captured
- 15 templates were denatured with 50 μ l of 100 Mm NaOH for 5 min. at ambient temperature, washed once with 50 μ l 50 Mm NH_4OH and three times with 100 μ l 10 Mm Tris/Cl, Ph 8.0. The single stranded DNA served as templates for PROBE reactions.

Primer Oligo Base Extension (PROBE) Reaction

- 20 The PROBE reaction was performed using Sequence 2.0 (USB Cat # E70775Z including buffer). dATP/DGTP and ddTTP were supplied by Boehringer-Mannheim (Cat # 1277049 and 1008382). d(CAG CTC TCA TTT TCC ATA C (SEQ ID. NO. 73) was used as PROBE primer (Figure 87). The following solutions were added tot he beads: 3.0 μ l H_2O , 1.0 μ l
- 25 reaction buffer, 1.0 μ l PROBE primer (10 pmol) and incubated at 65°C for 5 minutes followed by 37°C for 10 min. Then 0.5 μ l DTT, 3.5 μ l DNTPS/ddntp each 50 μ M and 0.5 μ l Sequenase (0.8 U) were added and incubated at 37°C for 10 min.

T4 Treatment of DNA

-192-

To generate blunt ended DNA, amplification products were treated with T4 DNA polymerase (Boehringer-Mannheim Cat# 1004786). The reactions were carried out according to the manufacturer's protocol for 20 min. at 11°C.

5 *Direct Size Determination of Extended Products*

To determine the size of the amplified product, MALDI-TOF was applied to one strand of the amplification product. samples were bound to beads, as described above, conditioned and denatured, as described below.

10 *DNA Conditioning*

After the PROBE reaction the supernatant was discarded and the beads were washed first in 50 μ l 700 mM NH_4 -citrate and second 50 μ l 50 mM NH_4 -citrate. The generated diagnostic products were removed for the template by heating the beads in 2 μ l H_2O at 80°C for 2 min.

15 The supernatant was used for MALDI-TOF analysis.

Sample Preparation and Analysis with MALDI-TOF Mass Spectrometry

Sample preparation was performed by mixing 0.6 μ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic citrate in 1:1 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$) with 0.3 μ l of diagnostic PROBE products in water on a sample target and allowed to air dry. Up to 100 samples were spotted on a probe target disk for introduction into the source region of an unmodified Perspective Voyager MALDI-TOF instrument operated in linear mode with delayed extraction and 5 and 30 kV on the target and conversion dynode, respectively. Theoretical average molecular mass ($M_r(\text{calc})$) of analytes were calculated from atomic compositions, reported experimental $M_r(M_r(\text{exp}))$ values are those of the singly-protonated form, determined using internal calibration with unextended primers in the case of PROBE reactions.

-193-

Conventional Analyses

Conventional analyses were performed by native polyacrylamide gel electrophoresis according to standard protocols. The diagnostic products were denatured with formamide prior to loading onto the gels and stained with ethidium bromide or silver, respectively.

RESULTS

The CKR-5 status of 10 randomly chosen DNA samples of healthy individuals were analyzed. Leukocyte DNA was amplified by PCR and an aliquot of the amplified product was analyzed by standard polyacrylamide gel electrophoresis and silver staining of the DNA (Figure 88). Four samples showed two bands presumably indicating heterozygosity for CKR-5, whereas the other 6 samples showed one band, corresponding to a homozygous gene (Figure 88). In the case where two bands were observed, they correspond to the expected size of 75 bp for the wildtype gene and 43 bp for the allele with the deletion (Figure 87). Where one band was observed, the size was about 75 bp which indicated a homozygous wildtype CKR-5 allele. One DNA sample derived from a presumably heterozygous one from a homozygous individual were used for all further analysis. To determine the molecular mass of the amplified product, DNA was subjected to matrix assisted laser desorption/ionization coupled with time of flight analysis (MALDI-TOF). Double stranded DNA, bound to streptavidin coated paramagnetic particles, was denatured and the strand released into the supernatant was analyzed. Figure 89A shows a spectrograph of a DNA sample, that was supposed to be heterozygous according to the result derived by polyacrylamide gel electrophoresis (Figure 88). The calculated mass of the sense strand for a wildtype gene is 23036 Da and for the sense strand carrying the deletion allele 13143 (Figure 87 and Table VI). Since many thermostable polymerases unspecifically add an adenosine to the

-194-

3' end of the product, those masses were also calculated. They are 23349 and 13456 Da. The masses of the observed peaks (Figure 89A) are 23119 Da, which corresponds to the calculated mass of a wildtype DNA strand where an adenosine has been added (23349 Da). Since no

5 peak with a mass of about 23036 Da was observed, the polymerase must have qualitatively added adenosine. Two peaks, which are close to each other, have a mass of 13451 and 13137 Da. This corresponds to the calculated masses of the allele, with the 32bp deletion. The higher mass peak corresponds to the product, where adenosine has been added

10 and the lower mass peak to the one without the unspecific adenosine. Both peaks have about the same height, indicating that to about half of the product adenosine has been added. The peak with a mass of 11682 Da is a doubly charged molecule of the DNA corresponding to 23319 Da ($2 \times 11682 \text{ Da} = 23364 \text{ Da}$). The peaks with masses of 6732 and

15 6575 Da are doubly charged molecules of the one with masses of 13451 and 13137 Da and the peak with 7794 Da corresponds to the triply charged molecule of 23319Da. Multiple charged molecules are routinely identified by calculation. Amplified DNA derived from a homozygous individual shows in the spectrograph (Figure 89C) one peak with a mass

20 23349.6 and a much smaller peak with a mass of 23039.9 Da. The higher mass peak corresponds to DNA resulting from a wildtype allele with an added adenosine, that has a calculated mass of 23349 Da. The lower mass peak corresponds to the same product without adenosine. Three further peaks with a mass of 11686, 7804.6 and 5852.5 Da

25 correspond to doubly, triply and quadruply charged molecules.

The unspecific added adenine can be removed from the amplified DNA by treatment of the DNA and T4 DNA polymerase. DNA derived from a heterozygous and a homozygous individual was analyzed after T4 DNA polymerase treatment. Figure 89B shows the spectrograph

-195-

derived from heterozygous DNA. The peak corresponding to the wildtype strand has a mass of 23008 Da indicating that the added adenine had been removed completely. The same is observed for the strand with a mass of 13140 Da.

- 5 The other three peaks are multiply charged molecules of the parent peaks. The mass spectrograph for the homozygous DNA shows one peak that has a mass of 23004 Da, corresponding to the wildtype DNA strand without an extra adenine added. All other peaks are derived from multiply charged molecules of this DNA. The amplified products can be
- 10 analyzed by direct determination of their masses, as described above, or by measuring the masses of products, that are derived from the amplified product in a further reaction. In this "primer oligo base extension (PROBE)" reaction, a primer that can be internal, as it is in the nested PCR, or identical to one of the PCR primers, is extended for just a few
- 15 bases before the termination nucleotide is incorporated. Depending on the extension length, the genotype can be specified. CKRΔ-F was used as a PROBE primer, and dATP/dGTP and ddTTP as nucleotides. The primer extension is AGT in case of a wildtype template and AT in case of the deletion (Figure 87). The corresponding masses are 6604 Da for the
- 20 wildtype and 6275 Da for the deletion, respectively. PROBE was applied to two standard DNAs. The spectrograph (Figure 90A) shows peaks with masses of 6604 Da corresponding to the wildtype DNA and at 6275 Da corresponding to the CKR-5 deletion allele (Table VIII). The peak at a mass of 5673 Da corresponds to CKRΔ-F (calculated mass of 5674 Da).
- 25 Further samples were analyzed in analogous way (Figure 90B). It is unambiguously identified as homozygous DNA, since the peak with a mass of 6607 Da corresponds to the wildtype allele and the peak with a mass of 5677 Da to the unextended primer. No further peaks were observed.

-196-

The example demonstrates that deletion analysis can be performed by mass spectrometry. As shown herein, the deletion can be analyzed by direct detection of single stranded amplified products, or by analysis of specifically generated diagnostic products (PROBE). In addition, as
 5 shown in the following Example 26, double stranded DNA amplified products can be analyzed.

10

Size	Calculated Mass	Measured Mass
wildtype w/o A	23036	23039/23009/23004
wildtype with A	23349	23319/23350
deletion w/o A	13143	13137/13139
deletion with A	13456	13451
PROBE		
wildtype	6604	6604/6608
deletion	6275	6275

15

All masses are in Dalton.

EXAMPLE 24

Pentaplex tc-PROBE

SUMMARY

20

The multiplexing of thermocycling primer oligo base extension (tc-PROBE) was performed using five polymorphic sites in three different apolipoprotein genes, which are thought to be involved in the pathogenesis of atherosclerosis. The apolipoprotein A IV gene (codons 347 and 360), the apolipoprotein E gene (codons 112 and 158), and the
 25 apolipoprotein B gene (codon 3500) were examined. All mass spectra were easy to interpret with respect to the five polymorphic sites.

-197-

MATERIALS AND METHODS

PCR Amplification

Human leukocytic genomic DNA was used for PCR. Listed below are the primers used for the separated amplification of portions of the

5 Apo A IV, Apo E and the Apo B genes:

Apo A IV:	A347F:	5'-CGA GGA GCT CAA GGC CAG AAT-3' (SEQ ID NO. 75)
	A360 R-2-bio:	*5'-CAG GGG CAG CTC AGC TCT C-3' (SEQ ID NO. 76)
Apo E:	ApoE-F:	5'-GGC ACG GCT GTC CAA GGA-3' (SEQ ID NO. 77)
	ApoE-R bio:	*5'-AGG CCG CGC TCG GCG CCC TC-3' (SEQ ID NO. 78)
10 Apo B:	ApoB-F2 bio:	*5'-CTT ACT TGA ATT CCA AGA GC-3' (SEQ ID NO. 79)
	Apo B-R:	5'-GGG CTG ACT TGC ATG GAC CGG A-3' (SEQ ID NO. 80)

* *biotinylated*

- Taq polymerase and 10x buffer were purchased from Boehringer-
15 Mannheim (Germany) and dNTPs for Pharmacia (Freiburg, Germany).
The total PCR reaction volume was 50 μ l including 10 pmol of each
primer and 10% DMSO (dimethylsulfoxide, Sigma) (no DMSO for the
PCR of the Apo B gene), with ~200 mg of genomic DNA used as
template and a final dNTP concentration of 200 μ M. Solutions were
20 heated to 80°C before the addition of 1U Taq polymerase; PCR
conditions were: 5 min at 95°C, followed by 2 cycles 30 sec 94°C, 30
sec 62°C, 30 sec 72°C, 2 cycles 30 sec 94°C 30 sec 58°C, 30 sec
72°C, 35 cycles of 30 sec at 94°C, 30 sec at 56°C, 30 sec at 72°C,
and a final extension time of 2 min at 72°C. To remove unincorporated
25 primers and nucleotides, amplified products were purified using the
"QIAquick" (Qiagen, Germany) kit, with elution of the purified products
in 50 μ L of TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 8.0).

-198-

Binding of the amplified product on beads

10 μ l of each purified amplified product was bound to 5 μ l DynaBeads (Dynal, M-280 Streptavidin) and denatured according to the protocol from Dynal. For the pentaplex tc-PROBE reaction the three
5 different amplified product (bound on the beads) were pooled.

Tc-PROBE

For the PROBE reaction the following primers were used:

- (Apo A) P347: 5'-AGC CAG GAC AAG-3' (SEQ ID NO. 81)
(Apo A) P360: 5'-ACA GCA GGA ACA GCA-3' (SEQ ID NO. 82)
10 (Apo E) P112: 5'-GCG GAC ATG GAG GAC GTG-3' (SEQ ID NO. 83)
(Apo E) P158: 5'-GAT GCC GAT GAC CTG CAG AAG-3' (SEQ ID NO. 84)
(Apo B) P3500: 5'-GTG CCC TGC AGC TTC ACT GAA GAC-3' (SEQ ID NO. 85)

- 15 The tc-PROBE was carried out in a final volume of 25 μ l containing 10 pmol of each primer listed above, 2.5 U Thermoquenase (Amersham), 2.5 μ L Thermoquenase buffer, and 50 μ M dTTP (final concentrations) and 200 μ M of ddA/C/GTP, respectively. Tubes containing the mixture were placed in a thermocycler and subjected to the following cycling
20 conditions: denaturation (94°C) the supernatant was carefully removed from the beads and 'desalted' by ethanol precipitation to exchange nonvolatile cations such as Na⁺ and K⁺ with NH₄⁺, which evaporated during the ionization process; 5 μ L 3M ammonium acetate (pH 6.5) 0.5 μ L glycogen (10 mg/mL, Sigma), 25 μ L H₂O, and 110 μ L absolute ethanol
25 were added to 25 μ L PROBE supernatant and incubated for 1 hour at 4°C. After a 10 min. centrifugation at 13,000 X g, the pellet was washed in 70% ethanol and resuspended in 1 μ L 18 Mohm/cm H₂O. A 0.35 μ L aliquot of resuspended DNA was mixed with 0.35 μ L matrix solution (0.7 M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium

-199-

citrate in 1:1 H₂O:CH₃CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using the Thermo Bioanalysis Version 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular masses (M_i(calc)) of the fragments were calculated from atomic compositions. External calibration generated from synthetic (ATCG)_n oligonucleotide (3.6-18kDa) was used. Positive ion spectra from 1-37500 Da were collected.

RESULTS

- Table VIII shows the calculated molecular masses of all possible extension products including the mass of the primer itself. Fig. 91 shows a respective MALDI-TOP MS spectra of a tc-PROBE using three different templates and 5 different PROBE primers simultaneously in ne reaction. Comparison of the observed and calculated masses (see table VIII) allows a fast genetic profiling of various polymorphic sites in an individual DNA sample. The sample presented in Figure 91 is homozygous for threonine and glutamine at position 347 and 360, respectively, in the apolipoprotein A IV gene, bears the epsilon 3 allele homozygous in the apolipoprotein E gene, and is also homozygous at the codon 3500 for arginine in the apolipoprotein B gene.

TABLE VIII

	SEQ ID	mass	allele
Apolipoprotein A IV			
5'-AGCCAGGACAAG-3' (347)	86	3688.40	unextended primer
5'-AGCCAGGACAAGTC-3'	87	4265.80	347Ser
5'-AGCCAGGACAAGA-3'	88	3985.60	347Thr
5'-ACAGCACCAACAGCA-3' (360)	89	4604.00	unextended primer
5'-ACAGCAGGAACAGCATC-3'	90	5181.40	360His

-200-

	SEQ ID	mass	allele
5'-ACAGCAGGAACAGCAG-3' (112)	91	4917.20	360Gln
Apolipoprotein E			
5'-GCGGACATGGAGGACGTG-3' (112)	92	5629.60	unextended primer
5'-GCGGACATGGAGGACGTGGC-3'	93	6247.00	112Cys
5'-GCGGACATGGAGGACGTGC-3'	94	5902.80	112Arg
5'-GATGCCGATGACCTGCAGAAG-3'(158)	95	6480.20	unextended primer
5'-GATGCCGATGACCTGCAGAAGC-3'	96	6753.40	158Arg
5'-GATGCCGATGACCTGCAGAAGTG-3'	97	7097.60	158Cys
Apolipoprotein B-100			
5'-GTGCCCTGCAGCTTCACTGAAGAC-3' (3500)	98	7313.80	unextended primer
5'-GTGCCCTGCAGCTTCACTGAAGACTG-3'	99	7931.20	3500Gln
5'-GTGCCCTGCAGCTTCACTGAAGACC-3'	100	7587.00	3500Arg

15

EXAMPLE 25**Sequencing Exons 5 to 8 of the p53 Gene by MALDI-TOF Mass Spectrometry****MATERIALS & METHODS**

20

Thirty-five cycles of PCR reactions were performed in a 96 well microliter plate with each well containing a total volume of 50 μ l including 200 ng genomic DNA, 1 unit Taq DNA polymerase, 1.5 mM Mg C1₂, 0.2mM dNTPx, 10 pmol of the forward primer and 6 or 8 of the biotinylated reverse primer. The sequences of PCR primers prepared

25

according to established chemistry (N.D. Sinha, J. Biernat, H. Kter, Tetrahed. Lett. 24:5843-5846 (1983) are as follows: exon 5:d(biotin-TATCTGTTCACCTTG TGCCC SEQ ID NO. 101) and d(biotin-CAGAGGCCTGGGGACCCTG SEQ ID NO.102); exon 6:

-201-

D(ACGACAGGGCTGGTTGCC SEQ ID NO. 103) and d(biotin-
ACTGACAACCACCCTTAAC SEQ ID NO. 104); exon 7:

d(CTGCTTGCCACAGGTCTC SEQ ID NO. 105) and d(biotin-
CACAGCAGGCCAGTGTGC SEQ ID NO. 106); exon 8:

- 5 d(GGACCTGATTTCTTACTG SEQ ID NO. 107) and d(biotin-
TGAATCTGAGGCATAACTG SEQ ID NO. 108).

- To each well of the 96-well microliter plate containing unpurified
amplified product, 0.1 mg of paramagnetic streptavidin beads (Dynal) in
10 μ l of 5 x binding solution (5 M NH_4OH) was added and incubated at 3
10 $^{\circ}\text{C}$ for 30 min.

Then beads were treated with 0.1 M NaOH at room temperature for 5
min followed by one wash with 50 mM NH_4OH at room temperature for
5 min followed by one wash with 50 mM Tris-HCl.

- Four dideoxy termination reactions were carried out in separate
15 wells of the microliter plate. A total of 84 reactions (21 primers x 4
reactions/primer) can be performed in a single microliter plate. To each
well containing immobilized single-stranded template, a total volume of 10
 μ l reaction mixture was added including 1x reaction buffer, 10 pmol of
sequencing primer, 250 mM of dNTPs, 25 mM of one of the ddNTPs,
20 and 1-2 units of Thermosequenase (Amersham). Sequencing reactions
were carried out on a thermal cycler using non-cycling conditions: 80 $^{\circ}\text{C}$,
1 min, 50 $^{\circ}\text{C}$, 1 min, 50 $^{\circ}\text{C}$ to 72 $^{\circ}\text{C}$, ramping 0.1 $^{\circ}/\text{sec}$, and 72 $^{\circ}\text{C}$, 5 min.
The beads were then washed with 0.7 M ammonium citrate followed by
0.05 M ammonium citrate. Sequencing products were then removed
25 from beads by heating the beads to 80 $^{\circ}\text{C}$ in 2 μ l of 50mM NH_4OH for 2
min. The supernatant was used for MALDI-TOF MS analysis.

Matrix was prepared as described in Kter, et al (Kter, H. *et al.*,
Nature Biotechnol. 14: 1123-1128 (1996)). This saturated matrix
solution was then diluted 1.52 times with pure water before use. 0.3 μ l

-202-

- of the diluted matrix solution was then diluted 1.52 times with pure water before use. 0.3 μ l of the diluted matrix solution was loaded onto the sample target and allowed to crystallize followed by addition of 0.3 μ l of the aqueous analyte. A Perseptive Voyager DE mass spectrometer
- 5 was used for the experiments, and the samples were typically analyzed in the manual mode. The target and middle plate were kept at +18.2 kV for 200 nanoseconds after each laser shot and then the target voltage was raised to +20kV. the ion guide wire in the flight tube was kept at -2V. Normally, 250 laser shots were accumulated for each sample. The
- 10 original spectrum was acquired under 500 MHz digitizing rate, and the final spectrum was smoothed by a 455 point average (Savitsky and Golay, (1964) Analytical Chemistry, 36:1627). Default calibration of the mass spectrometer was used to identify each peak and assign sequences. The theoretical mass values of two sequencing peaks were
- 15 used to recalibrate each spectrum. (D.P. Little, T.J. Cornish, M.J. O'Donnel, A. Braun, R.J. Cotter, H. Kter, Anal. Chem., submitted).

RESULTS

- Alterations of the p53 gene are considered to be a critical step in the development of many human cancers (Greenblatt, et al., (1994)
- 20 Cancer Res. 54, 4855-4878; C.C. Harris, (1996) J. Cancer, 73, 261-269; and D. Sidransky and M. Hollstein, (1996) Annu.Res.Med., 47,285-301). Mutations may serve as molecular indicators of clonality or as early markers of relapse in a patient with a previously identified mutation in a primary tumor (Hainaut, et al., (1997) Nucleic Acid Res., 25, 151-
- 25 157). The prognosis of the cancer may differ according to the nature of the p53 mutations present (H.S. Goh et al., (1995) Cancer Res, 55, 5217-5221). Since the discovery of the p53 gene, more than 6000 different mutations have been detected. Exons 5-8 were selected as

-203-

sequencing targets where most of the mutations cluster (Hainaut et al. (1997) *Nucleic Acids Res.*, 25, 151-7).

Figure 96 schematically depicts the single tube process for target amplification and sequencing, which was performed, as described in
5 detail in the Materials and Methods. Each of exon 5-8 of the p53 gene was PCR amplified using flanking primers in the intron region; the downstream primer was biotinylated. Amplifications of different exons were optimized to use the same cycling profile, and the products were used without further purification. PCR reactions were performed in a 96 well
10 microliter plate and the product generated in one well was used as the template for one sequencing reaction. Streptavidin-coated magnetic beads were added to the same microliter plate and amplified products were immobilized. The beads were then treated with NaOH to generate immobilized single-stranded DNA as sequencing template. The beads
15 were washed extensively with Tris buffer since remaining base would reduce the activity of sequencing enzyme.

A total of 21 primers were selected to sequence exon 5-8 of the p53 gene by primer walking. The 3'-end nucleotide of all the primers is located at the site where no known mutation exists. Four termination
20 reactions were performed separately which resulted in a total of 84 sequencing reactions on the same PCR microliter plate. Non-cycling conditions were adopted for sequencing since streptavidin coated beads do not tolerate the repeated application of high temperature. Sequencing reactions were designed so that mt terminated fragments were under 70
25 nucleotides, a size range easily accessible by MALDI-TOF MS and yet long enough to sequence through the next primer binding site. Thermequenase was the enzyme of choice since it could reproducibly generate a high yield of sequencing products in the desired mass range. After the sequencing reactions, the beads were washed with ammonium

-204-

ion buffers to replace all other cations. The sequencing ladders were then removed from the beads by heating in ammonium hydroxide solution or simply in water.

5 A sub-microliter aliquot of each of the 84 sequencing reactions was loaded onto one MS sample holder containing preloaded matrix. Figure 94 gives an example of sequencing data generated from one primer; four spectra are superimposed.

10 All sequencing peaks were well resolved in the mass range needed to read through the next sequencing primer site. Sometimes doubly charged peaks were observed which could be easily identified by correlating the mass to that of the singly charged ion. False stops generated by early termination of the enzymatic extension can be observed cle to the primer site. Since the mass resolution is high enough, it is easy to differentiate the false stop peaks from the real
15 sequencing peaks by calculating the mass difference of the neighboring peaks and crs comparing the four spectra. Additionally, mt primers generated detectable data through the region of the downstream primer binding site thereby covering the false stop region.

20 Using optimized procedures of amplification, sequencing, and conditioning, exons 5-8 of the p53 gene were successfully sequenced. Correct wildtype sequence data were obtained from all exons with a mass resolution about 300 to 800 over the entire mass range. The overall mass accuracy is 0.05% or better. The average amount of each sequencing fragment loaded on the MS sample holder is estimated to be
25 50 fmol or less.

This example demonstrates the feasibility of sequencing exons of a human gene by MALDI-TOF MS. Compare to gel-based automated fluorescent DNA sequencing, the read lengths are shorter. Microchip technology can be incorporated to provide for parallel processing.

-205-

Sequencing products generated in the microtiter plate can be directly transferred to a microchip which serves as a launching pad for MALDI-TOF MS analysis. Robot-driven serial and parallel nanoliter dispensing tools are being used to produce 100-1000 element DNA arrays on <1"

- 5 square chips with flat or geometrically altered (e.g., with wells) surfaces for rapid mass spectrometric analysis.

- Figure 94 shows an MS spectrum obtained on a chip where the sample was transferred from a microtiter plate by a pintool. The estimated amount of each termination product loaded is 5 fmol or less
10 which is in the range of amounts used in conventional Sanger sequencing with radiolabeled or fluorescent detection (0.5-1 fmol per fragment). The low volume MALDI sample deposition has the advantages of miniaturization (reduced reagent cts), enhanced reproducibility and automated signal acquisition.

15 **EXAMPLE 26**

Direct detection of synthetic and biologically generated double-stranded DNA by MALDI-TOF MS

Introduction

- Typically, matrix-associated laser desorption/ionization (Karas, et.
20 al., (1989) *Int. J. Mass Spectrom, Ion Processes*, 92, 231) time-of-flight mass spectrometry (MALDI-TOF MS) of DNA molecules which are double stranded (ds) in solution yields molecular ions representative of the two single stranded components (Tang, *et al.* (1994) Rapid Commun. Mass Spectrom. 8:183; Tang, *et al.* (1995) Nucleic Acids Res. 23:3126;
25 Benner, *et al.* (1995) Rapid Commun. Mass Spectrom. 9:537; Liu, *et al.* (1995) Anal. Chem. 67:3482; Siegert *et al.* (1996) Anal. Biochem. 243:55; and Doktycz, *et al.* (1995) Anal. Biochem. 230:205); this has been observed in several reports dealing with biologically generated DNA from a polymerase chain reaction (PCR) amplification (Tang, *et al.* (1994)

-206-

- Rapid Commun. Mass Spectrom. 8:183; Liu, *et al.* (1995) Anal. Chem. 67:3482; Siegert *et al.* (1996) Anal. Biochem. 243:55; and Doktycz, *et al.* (1995) Anal. Biochem. 230:205). It is not clear whether the double strand is destabilized because of the decreased pH in the matrix environment or because of absorbance by the duplex during desorption/ionization/acceleration of an energy sufficient to overcome the attractive van der Waals and "stacking" stabilization forces (Cantor and Shimmel, Biophysical Chemistry Part I: The conformation of Biomolecules, W.H. Freeman, New York, (1980), 176). When analyte is present at high concentrations formation of non-specific gas-phase DNA multimers is, as with proteins (Karas, *et. al.*, (1989) Int. J. Mass Spectrom, Ion Processes 92:231), common; however, Lecchi and Pannell (Lecchi *et al.* (1995) J. Am. Soc. Mass Spectrom. 6:972) have provided strong evidence for *specific* Watson Crick (WC) base pairing being maintained in the gas phase. They detected these specific dimers when using 6-aza-2-thiothymine as a matrix, but did not observe them with 3-hydroxypiccolinic acid (3-HPA) or 2,4,6- hydroxyacetophenone matrix. As described below, by using a low acceleration voltage of the ions and preparing samples for MALDI analysis at reduced temperatures, routine detection of dsDNA is possible.

MATERIALS AND METHODS

- Synthetic DNA.* Oligonucleotides were synthesized (Sinha, *et al.* (1984) Nucleic Acids Res., 12, 4539) on a Perspective Expedite DNA synthesizer and reverse phase HPLC purified in-house. Sequences were:
- 50-mer (15337 Da): 5'-TTG CGT ACA CAC TGG CCG TCG TTT TAC AAC GTC GTG ACT GGG AAA ACC CT-3' (SEQ ID NO. 109); 27-mer_c (complementary, 8343 Da): 5'-GTA AAA CGA CGG CCA GTG TGT ACG CAA-3' (SEQ ID NO. 110); 27-mer_{nc} (non-complementary, 8293 Da): 5'-TAC TGG AAG GCG ATC TCA GCA ATC AGC-3' (SEQ ID NO. 111).

-207-

100 μ M stock solutions were diluted to 20, 10, 5, and 2.5 μ M using 18Mohm/cm H₂O. 2 μ L each of equimolar solutions of the 50-mer and either 27-mer_c or 27-mer_{nc} were mixed and allowed to anneal at room temperature for 10 minutes. 0.5 μ L of these mixtures were mixed directly
5 on a sample target with 1 μ L matrix (0.7 M 3-HPA, 0.07 M ammonium citrate in 50% acetonitrile) and allowed to air dry.

Biological DNA. Enzymatic digestion of human genomic DNA from leukocytes was performed. PCR primers (forward, 5'-GGC ACG GCT GTC CAA GGA G-3' (SEQ ID NO. 112)); reverse, 5'-AGG CCG CGC TCG
10 GCG CCC TC-3' (SEQ ID NO. 113) to amplify a portion of exon 4 of the apolipoprotein E gene were delineated from the published sequence (Das et al., (1985) *J. Biol. Chem.*, 260 6240). Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs from Pharmacia (Freiburg, Germany). The total reaction volume was 50
15 μ L including 20 pmol of each primer and 10% DMSO (dimethylsulfoxide, Sigma) with approximately 200 ng of genomic DNA used as template. Solutions were heated to 80°C before the addition of IU polymerase; PCR conditions were: 2 min at 94°C, followed by 40 cycles of 30 sec at 94°C, 45 sec at 63°C, 30 sec at 72°C, and a final extension time of 2
20 min at 72°C. While no quantitative data was collected to determine the final yield of amplified product, it is estimated that -2pmol were available for the enzymatic digestion.

CfoI and RsaI and reaction buffer L were purchased from Boehringer-Mannheim. 20 μ L of amplified products were diluted with 15 μ L
25 water and 4 μ L buffer L; after addition of 10 units of restriction enzymes the samples were incubated for 60 min at 37°C. For precipitation of digest products 5 μ L of 3M ammonium acetate (pH 6.5), (5 μ L glycogen (Braun, et al. (1997) *Clin. Chem.* 43:1151) (10mg/ml, Sigma), and 110 μ L absolute ethanol were added to 50 μ L of the analyte solutions and stored

-208-

for 1 hour at room temperature. After at 10 min centrifugation at 13,000 X g, the pellet was washed in 70% ethanol and resuspended in 1 μ l 18Mohm/cm H₂O.

- Sample preparation and analysis by MALDI-TOF MS.* 0.35 μ l of
- 5 resuspended DNA was mixed with 0.35-1.3 μ L matrix solution (0.7M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium citrate in 1:1 H₂O:CH₃CN) (Wu, et al. (1993) Rapid Commun. Mass Spectrom. 7:142) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-
- 10 TOF instrument operated in pitive ion reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular masses ($M_r(\text{calc})$) of the fragments were calculated from atomic compositions; the mass of a proton (1.08 Da) was subtracted from raw data values in reporting experimental molecular masses
- 15 ($M_r(\text{exp})$) as neutral basis. External calibration generated from eight peaks (2000-18000 Da) was used for all spectra.

Results and Discussion

- Figure 96A is a MALDI-TOF mass spectrum of a mixture of the synthetic 50-mer with (non-complementary) 27-mer_{nc} (each 10 μ M, the
- 20 highest final concentration used in this study); the laser power was adjusted to just above the threshold irradiation for ionization. The peaks at 8.30 and 15.34 kDa represent singly charged ions derived from the 27- and 50-mer single strands, respectively. Poorly resolved low intensity signals at -16.6 and -30.7 kDa represent homodimers of 27-
- 25 and 50-mer, respectively; that at 23.6 kDa is consistent with a heterodimer containing one 27-mer and one 50-mer strand. Thus low intensity dimer ions representing all possible combinations from the two non-complementary oligonucleotides (27 + 27; 27 + 50; 50 + 50) were observed. Increasing the irradiance even to a point where depurination

-209-

peaks dominated the spectrum resulted in slightly higher intensities of these dimer peaks. Note that the hybridization was performed at room temperature and with a very low salt concentration, conditions at which non-specific hybridization may occur.

- 5 Figure 96 shows a MALDI-TOF spectrum of the same 50-mer mixed with (complementary) 27-mer_c; the final concentration of each oligonucleotide was again 10 μ M. Using the same laser power as in Figure 96A, intense signals were again observed at 88.34 and 15.34 Kda, consistent with single stranded 27- and 50-mer, respectively.
- 10 Homodimer peaks (27 + 27; 50 + 50) were barely apparent in the noise; however, singly (23.68 Kda) and doubly (11.84k Da) charged heterodimer (27 + 50) peaks were dominant. Although the 23.68 Kda dimer peak could be detected from all irradiated positions, its intensity relative to the monomer peaks varied slightly from spot-to-spot.
- 15 Repeating the experiment with individual oligonucleotide concentrations of 5, 2.5, and 1.25 μ M resulted in decreasing amounts of the 27-/50-mer Watson-Crick dimer peak relative to the 27- and 50-mer single stranded peaks. At the lowest concentrations, the observation of dimer was "crystal-dependent", that is, irradiation of some crystals produced
- 20 significant 27-/50-mer dimer signal, while other crystals reproducibly yielded very little or none. This indicates that the incorporation of dsdna into the matrix crystals or the effectiveness of retaining this interaction through the ionization/desorption process is dependent upon the microscopic properties of the crystals, and/or that there exist steep
- 25 concentration gradients of the duplex throughout the sample.

Thus the Figure 96 spectra provide strong evidence that specific WC base paired dsdna can be observed using gentle laser conditions with high concentrations of oligonucleotides in this mass range, the first report of this using a 3-HPA matrix. The study was extended to a

-210-

complex mixture of dsdna derived from an enzymatic digest (Rsal/Cfol) of a region of exon 4 of the apolipoprotein E gene (Das et. al., (1985) *J. Biol. Chem.*, 260 6240); expected fragment masses are given in Table IX.

5

Table IX
Cfol/Rsal Digestion Products from ApoE gene exon 4²

	bases ^b		ssDNA		(Da)		dsdna (Da)
	(+)	(-)	(+)	(-)	(+)	(-)	
10	11	13	3428		4025		7453
	16		5004		4924		9928
	18		5412		5750		11162
	17	19	5283		5880		11163
	19		5999		5781		11780
15	24	22	7510		6745		14225
	31	29	9628		9185		18813
	36	38	11279		11627		22906
	48		14845		14858		29703
	55	53	17175		16240		33415

20 ^aε3 allele has no 17/19 or 19/19 pairs; ε4 allele contains no 36/38 pair.

^b(+) sense strand, (-) antisense strand

After the digestion step, the samples were purified and concentrated by ethanol precipitation and resuspended in 1μL H₂O before mixing them at room temperature with matrix on the sample target. Nearly 20 peaks

25 ranging in mass from 3.4-17.2 Kda were resolved in the products' MALDI spectrum (Figure 97A), all consistent with denatured single stranded components of the double strand (Table IX). Many such analyses of similar biological products over a period of months also yielded spectra with negligible dsdna, consistent with previous reports

30 (Tang, *et al.* (1994) *Rapid Commun. Mass Spectrom.* 8:183; Liu, *et al.*

-211-

(1995) Anal. Chem. 67:3482; Siegert *et al.* (1996) Anal. Biochem. 243:55; and Doktycz, *et al.* (1995) Anal. Biochem. 230:205); contrarily, intact double strands were observed under similar conditions for the synthetic DNA (Figure 96A). It is difficult to estimate the strand

- 5 concentration available after the biological reactions, but presumably that it was far lower than that at which dimerization of synthetic samples occurred. Furthermore, maintaining specific hybrids within the two-component synthetic mixture may be kinetically favored relative to the far more complex mixture of 20 single-stranded DNA components from
- 10 the digest.

- The effect of reduced temperature on maintaining dsDNA was tested. An aliquot of the digested DNA solution, the matrix, pipette, pipette tips, and the stainless steel sample target were stored in a 4°C "cold room" for 15 minutes; as with normal preparations matrix, and
- 15 then analyte, were spotted on the target and allowed to co-crystallize while air drying. Crystallization for mixtures of 300 nL 3HPA (50% acetonitrile) with 300 nL analyte required ~1 minute at room temperature but ~15 minutes at the reduced temperature. Sample spots prepared in the cold room environment typically contained a high
- 20 proportion of large transparent crystals.

- MALDI-TOF analysis of an ApoE digest aliquot prepared at reduced temperature produced the Figure 97B spectrum. While the low mass range appeared qualitatively similar to Figure 97A, dramatic differences above 8 kDa were observed. Only signals consistent with single strands
- 25 (Table IX) were observed in Figure 97A, but the Figure 97B cold room prepared samples did not yield signals for the same masses except below 8 kDa. Even more striking were the additional high mass peaks in Figure 97B; clearly these represent dimer peaks containing lower mass components. As was done with the synthetic DNA, it was important to

-212-

determine whether these represent non-specific heterodimers, specific WC heterodimers, or nonspecific homodimers. Consider first the 33.35 kDa fragment. Ignoring the unlikely possibility that the high mass fragment represents a trimer or higher multimer, as a dimer it must only

5 contain the highest mass ssDNA components, i.e., the > 16 kDa. Homodimerization of the 15.24 and 17.18 kDa fragments would result in 32.49 and 34.35 kDa peaks, respectively; corresponding mass errors for these incorrect assignments relative to the observed 33.35 kDa would be -2.6% and + 3.0% respectively. A far better match is achieved if this

10 peak originates from a heterodimer of the two highest mass single stranded fragments; their summed mass ($16.24 + 17.18 = 33.42$ kDa) differed by 0.2% from the observed dimer mass 33.35 kDa, an acceptable mass error for MALDI-TOF analysis of large DNA fragments using external calibration. Likewise, the 29.66 kDa fragment was

15 measured only 0.13% lower than the 29.70 Da expected for a heterodimer of 48-mers; the sum of no other possible homodimers or heterodimers were within a reasonable range of this mass. Similar arguments could be made for the 22.89 and 18.83 kDa fragments, representing 36-/38-mer and 31-/29-heterodimers, respectively; the

20 signal at 14.86 kDa is consistent with singly charged single stranded and doubly charged double-stranded 48-mer. The agreement of the Figure 97B masses above 15 kDa with the of dsDNA expected from this digest and the absence of homodimers and non-specific heterodimers at random masses indicated that the base pairings were indeed highly specific and

25 provided further evidence that gas-phase WC interactions may be retained in MALDI-generated ions.

Figure 98 shows a MALDI-TOF spectrum of an $\epsilon 4$ allele, which, unlike the $\epsilon 3$, was expected to yield no 36-/38-mer pair upon CfoI/RsaI digestion. The $\epsilon 3$ and $\epsilon 4$ mass spectra were similar except that

-213-

abundant 22.89 kDa fragment in Figure 97B was not present in Figure 98; with this information alone (Table IX) $\epsilon 3$ and $\epsilon 4$ alleles were easily distinguished, thereby demonstrating the genotyping by direct measurement of dsDNA by MALDI-TOF MS. Similarly dsDNA could be

5 ionized, transferred to the gas phase, and detected by MALDI-TOF MS. The acceleration voltage typically employed on our instrument was only 5kV corresponding to 1.5kV/mm up to -2 mm from the sample target, with the electric field strength decreasing rapidly with distance from the sample target. Most previous work used at least 20kV acceleration

10 (Lecchi *et al.* (1995) J. Am. Soc. Mass Spectrom. 6:972); in one exception a 27-mer dsDNA was detected using a frozen matrix solution and 100 V acceleration (Nelson, *et al.* (1990) Rapid Commun. Mass Spectrom. 4:348). Without being bound by any theory MALDI-induced "denaturation" of dsDNA may be due to gas-phase collisional activation

15 that disrupts the WC pairing when high acceleration fields are employed, analogous to the denaturation presumed to be a first step in the fragmentation used for sequencing the single stranded components of dsDNA using electrospray ionization (McLafferty *et al.* (1996) Int. J. Mass Spectrom., Ion Processes). It appears that the high salt

20 concentrations (typically >10mM NaCl or KCl) required to stabilize WC paired dsDNA in solution are unsuitable for MALDI analysis (Nordhoff *et al.* (1993) Nucleic Acids Res. 21:3347); reducing the concentration of such non-volatile cations is necessary to avoid cation-adducted MALDI signals, but destabilizes the double strands in solution. The low pH

25 conditions of the matrix environment should also destabilize the duplex. As shown in Figures 97B and 98, storing and preparing even low concentrations of the biological samples at reduced temperature at least in part offset these denaturing effects, especially for longer strands where melting temperatures are higher due to a more extensive hydrogen

-214-

bonding network. The conditions used here are recognized to be very non-stringent annealing conditions.

The low mass tails on high mass dsDNA peaks (e.g., Figure 97B, 232 kDa) are consistent with depurination generated to a higher extent
5 than the sum of depurination from each of the single strands combined. Although depurination in solution is an acid-catalyzed reaction, the weakly acidic conditions in the 3-HPA matrix do not induce significant depurination; molecular ion signals from a mixed-base 50-mer measured with De-MALDI-TOF had only minor contributions from depurination
10 peaks (Juhaz, *et al.* (1996) Anal. Chem. 68:941). Depurination from the single stranded components of the gas-phase dsDNA is observed even though these bases are expected to be hydrogen bonded to the complementary base of the accompanying strand, implying that covalent bonds are being broken before the strand is denatured.

15

EXAMPLE 27

Efficiency and Specificity Assay for Base-Specific Ribonucleases

Aliquots sampled at regular time intervals during digestion of selected synthetic 20 to 25 mers were analyzed by mass spectrometry. Three of the RNAses were found to be efficient and specific. These
20 include: the G-specific T₁, the A-specific U₂ and the A/U-specific PhyM. The ribonucleases presumed to be C-specific were found to be less reliable, e.g., did not cleave at every C or also cleaved at U in an unpredictable manner. The three promising RNAses all yielded cleavage at all of the predicted positions and a complete sequence coverage was
25 obtained. In addition, the presence of cleavage products containing one or several uncleaved positions (short incubation times), allowed alignment of the cleavage products. An example of the MALDI-spectrum of an aliquot sampled after T₁ digest of a synthetic 20-mer [SEQ ID NO:114] RNA is shown in Figure 100.

-215-

EXAMPLE 28**Immobilization of amplified DNA targets to silicon wafers****Silicon surface preparation**

- Silicon wafers were washed with ethanol, flamed over bunsen
- 5 burner, and immersed in an anhydrous solution of 25% (by volume) 3-aminopropyltriethoxysilane in toluene for 3 hours. The silane solution was then removed, and the wafers were washed three times with toluene and three times with dimethyl sulfoxide (DMSO). The wafers were then incubated in a 10mM anhydrous solution of N-succinimidyl (4-
- 10 iodoacetyl) aminobenzoate (SIAB) (Pierce Chemical, Rockford, IL) in anhydrous DMSO. Following the reaction, the SIAB solution was removed, and the wafers were washed three times with DMSO. In all cases, the iodoacetamido-functionalized wafers were used immediately to minimize hydrolysis of the labile iodoacetamido-functionality.
- 15 Additionally, all further wafer manipulations were performed in the dark since the iodoacetamido-functionality is light sensitive.

Immobilization of amplified thiol-containing nucleic acids

- The SIAB-conjugated silicon wafers were used to analyze specific free thiol-containing DNA fragments of a particular amplified DNA target
- 20 sequence. A 23-mer oligodeoxynucleotide containing a 5'-disulfide linkage [purchased from Operon Technologies; SEQ ID NO: 117] that is complementary to the 3'-region of a 112 bp human genomic DNA template [Genebank Acc. No.: Z52259; SEQ ID NO: 118] was used as a primer in conjunction with a commercially available 49-mer primer, which
- 25 is complementary to a portion of the 5'-end of the genomic DNA [purchased from Operon Technologies; SEQ ID NO: 119], in PCR reactions to amplify a 135 bp DNA product containing a 5'-disulfide linkage attached to only one strand of the DNA duplex [SEQ ID NO: 120].

-216-

The PCR amplification reactions were performed using the Amplitaq GoldKit [Perkin Elmer Catalog No. N808-0249]. Briefly, 200 ng 112 bp human genomic DNA template was incubated with 10 μ M of 23-mer primer and 8 μ M of commercially available 49-mer primer, 10 mM dNTPs, 1 unit of Amplitaq Gold DNA polymerase in the buffer provided by the manufacturer and PCR was performed in a thermocycler.

The 5'-disulfide bond of the resulting amplified product was fully reduced using 10 mM tris-(2-carboxyethyl) phosphine (TCEP) (Pierce Chemical, Rockford, IL) to generate a free 5'-thiol group. Disulfide reduction of the modified oligonucleotide was monitored by observing a shift in retention time on reverse-phase FPLC. It was determined that after five hours in the presence of 10 mM TCEP, the disulfide was fully reduced to a free thiol. Immediately following disulfide cleavage, the modified oligonucleotide was incubated with the iodacetamido-functionalized wafers and conjugated to the surface of the silicon wafer through the SIAB linker. To ensure complete thiol deprotonation, the coupling reaction was performed at pH 8.0. Using 10mM TCEP to cleave the disulfide and the other reaction conditions described above, it was possible to reproducibly yield a surface density of 250 fmol per square mm of surface.

Hybridization and MALDI-TOF Mass spectrometry

The silicon wafer conjugated with the 135 bp thiol-containing DNA was incubated with a complementary 12-mer oligonucleotide [SEQ ID NO: 121] and specifically hybridized DNA fragments were detected using MALDI-TOF MS analysis. The mass spectrum revealed a signal with an observed experimental mass-to-charge ratio of 3618.33; the theoretical mass-to-charge ratio of the 12-mer oligomer sequence is 3622.4 Da.

Thus, specific DNA target molecule that contain a 5'-disulfide linkage can be amplified. The molecules are immobilized at a high

-217-

density on a SIAB-derivatized silicon wafer using the methods described herein and specific complementary oligonucleotides may be hybridized to these target molecules and detected using MALDI-TOF MS analysis.

EXAMPLE 29

5 Use of High Density Nucleic Acid Immobilization to Generate Nucleic Acid Arrays

Employing the high density attachment procedure described in EXAMPLE 28, an array of DNA oligomers amenable to MALDI-TOF mass spectrometry analysis was created on a silicon wafer having a plurality of
10 locations, e.g., depressions or patches, on its surface. To generate the array, a free thiol-containing oligonucleotide primer was immobilized only at the selected locations of the wafer [e.g., see EXAMPLE 28]. The each location of the array contained one of three different oligomers. To
15 demonstrate that the different immobilized oligomers could be separately detected and distinguished, three distinct oligonucleotides of differing lengths that are complementary to one of the three oligomers were hybridized to the array on the wafer and analyzed by MALDI-TOF mass spectrometry.

Oligodeoxynucleotides

20 Three sets of complementary oligodeoxynucleotide pairs were synthesized in which one member of the complementary oligonucleotide pair contains a 3'- or 5'-disulfide linkage [purchased from Operon Technologies or Oligos, Etc.]. For example, Oligomer 1 [d(CTGATGCGTCGGATCATCTTTTTT-SS); SEQ ID NO: 122] contains a
25 3'-disulfide linkage whereas Oligomer 2 [d(SS-CCTCTTGGGAAGTGTAGTATT); a 5'-disulfide derivative of SEQ ID NO: 117] and Oligomer 3 [d(SS-GAATTCGAGCTCGGTACCCGG); a 5'-disulfide derivative of SEQ ID NO: 115] each contain a 5'-disulfide linkage.

-218-

The oligonucleotides complementary to Oligomers 1-3 were designed to be of different lengths that are easily resolvable from one another during MALDI-TOF MS analysis. For example, a 23-mer oligonucleotide [SEQ ID NO: 123] was synthesized complementary to a portion of Oligomer 1, a 12-mer oligonucleotide [SEQ ID NO: 121] was synthesized complementary to a portion of Oligomer 2 and a 21-mer [SEQ ID NO: 116] was synthesized complementary to a portion of Oligomer 3. In addition, a fourth 29-mer oligonucleotide [SEQ ID NO: 124] was synthesized that lacks complementarity to any of the three oligomers. This fourth oligonucleotide was used as a negative control.

Silicon surface chemistry and DNA immobilization

(a) 4 x 4 (16-location) array

A 2 X 2 cm² silicon wafer having 256 individual depressions or wells in the form of a 16 X 16 well array was purchased from a commercial supplier [Accelerator Technology Corp., College Station, Texas]. The wells were 800 X 800 μm^2 , 120 μm deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in iodoacetamido functionalities on the surface [e.g., see EXAMPLE 28].

To prepare the oligomers for coupling to the various locations of the silicon array, the disulfide bond of each oligomer was fully reduced using 10 mM TCEP as depicted in EXAMPLE 28, and the DNA resuspended at a final concentration of 10 μM in a solution of 100 mM phosphate buffer, pH 8.0. Immediately following disulfide bond reduction, the free-thiol group of the oligomer was coupled to the iodoacetamido functionality at 16 locations on the wafer using the probe coupling conditions essentially as described above in EXAMPLE 28. To accomplish the separate coupling at 16 distinct locations of the wafer,

-219-

the entire surface of the wafer was not flushed with an oligonucleotide solution but, instead, an ~30-nl aliquot of a predetermined modified oligomer was added in parallel to each of 16 locations (i.e., depressions) of the 256 wells on the wafer to create a 4 x 4 array of immobilized DNA
5 using a robotic pintool.

The robotic pintool consists of 16 probes housed in a probe block and mounted on an X Y, Z robotic stage. The robotic stage was a gantry system which enables the placement of sample trays below the arms of the robot. The gantry unit itself is composed of X and Y arms which
10 move 250 and 400 mm, respectively, guided by brushless linear servo motors with positional feedback provided by linear optical encoders. A lead screw driven Z axis (50 mm vertical travel) is mounted to the xy axis slide of the gantry unit and is controlled by an in-line rotary servo motor with positional feedback by a motor-mounted rotary optical encoder. The
15 work area of the system is equipped with a slide-out tooling plate that holds five microtiter plates (most often, 2 plates of wash solution and 3 plates of sample for a maximum of 1152 different oligonucleotide solutions) and up to ten 20x20 mm wafers. The wafers are placed precisely in the plate against two banking pins and held secure by
20 vacuum. The entire system is enclosed in plexi-glass housing for safety and mounted onto a steel support frame for thermal and vibrational damping. Motion control is accomplished by employing a commercial motion controller which was a 3-axis servo controller and is integrated to a computer; programming code for specific applications is written as
25 needed.

To create the DNA array, a pintool with assemblies that have solid pin elements was dipped into 16 wells of a multi-well DNA source plate containing solutions of Oligomers 1-3 to wet the distal ends of the pins, the robotic assembly moves the pin assembly to the silicon wafer, and

-220-

the sample spotted by surface contact. Thus, one of modified Oligomers 1-3 was covalently immobilized to each of 16 separate wells of the 256 wells on the silicon wafer thereby creating a 4 x 4 array of immobilized DNA.

- 5 In carrying out the hybridization reaction, the three complementary oligonucleotides and the negative control oligonucleotide were mixed at a final concentration of 10 μ M for each oligonucleotide in 1 ml of TE buffer [10 mM Tris-HCl, pH 8.0, 1 mM EDTA] supplemented with 1 M NaCl, and the solution was heated at 65°C for 10 min. Immediately thereafter,
- 10 the entire surface of the silicon wafer was flushed with 800 μ l of the heated oligonucleotide solution. The complementary oligonucleotides were annealed to the immobilized oligomers by incubating the silicon array at ambient temperature for 1 hr, followed by incubation at 4°C for at least 10 min. Alternatively, the oligonucleotide solution can be added
- 15 to the wafer which is then heated and allowed to cool for hybridization.

- The hybridized array was then washed with a solution of 50 mM ammonium citrate buffer for cation exchange to remove sodium and potassium ions on the DNA backbone (Pieles *et al.*, (1993) Nucl. Acids Res. 21:3191-3196). A 6-ml aliquot of a matrix solution of 3-
- 20 hydroxypicolinic acid [0.7 M 3-hydroxypicolinic acid-10 % ammonium citrate in 50 % acetonitrile; see Wu *et al.* Rapid Commun. Mass Spectrom. 7:142-146 (1993)] was added in series to each location of the array using a robotic piezoelectric serial dispenser (*i.e.*, a piezoelectric pipette system).

- 25 The piezoelectric pipette system is built on a system purchased from Microdrop GmbH, Norderstedt Germany and contains a piezoelectric element driver which sends a pulsed signal to a piezoelectric element bonded to and surrounding a glass capillary which holds the solution to be dispensed; a pressure transducer to load (by negative pressure) or

-221-

empty (by positive pressure) the capillary; a robotic xyz stage and robot driver to maneuver the capillary for loading, unloading, dispensing, and cleaning, a stroboscope and driver pulsed at the frequency of the piezo element to enable viewing of 'suspended' droplet characteristics;

- 5 separate stages for source and designation plates or sample targets (i.e. Si chip); a camera mounted to the robotic arm to view loading to designation plate; and a data station which controls the pressure unit, xyz robot, and piezoelectric driver.

- The 3-HPA solution was allowed to dry at ambient temperature
10 and thereafter a 6-nl aliquot of water was added to each location using the piezoelectric pipette to resuspend the dried matrix-DNA complex, such that upon drying at ambient temperature the matrix-DNA complex forms a uniform crystalline surface on the bottom surface of each location.

15 **MALDI-TOF MS analysis**

- The MALDI-TOF MS analysis was performed in series on each of the 16 locations of the hybridization array illustrated in Figure 6 essentially as described in EXAMPLE 28. The resulting mass spectrum of oligonucleotides that specifically hybridized to each of the 16 locations of
20 the DNA hybridization revealed a specific signal at each location representative of observed experimental mass-to-charge ratio corresponding to the specific complementary nucleotide sequence.

- For example, in the locations that have only Oligomer 1 conjugated thereto, the mass spectrum revealed a predominate signal with an
25 observed experimental mass-to-charge ratio of 7072.4 approximately equal to that of the 23-mer; the theoretical mass-to-charge ratio of the 23-mer is 7072.6 Da. Similarly, specific hybridization of the 12-mer oligonucleotide to the array, observed experimental mass-to-charge ratio of 3618.33 Da (theoretical 3622.4 Da), was detected only at those

-222-

locations conjugated with Oligomer 2 whereas specific hybridization of MJM6 (observed experimental mass-to-charge ratio of 6415.4) was detected only at those locations of the array conjugated with Oligomer 3 [theoretical 6407.2 Da].

- 5 None of the locations of the array revealed a signal that corresponds to the negative control 29-mer oligonucleotide (theoretical mass-to-charge ratio of 8974.8) indicating that specific target DNA molecules can be hybridized to oligomers covalently immobilized to specific locations on the surface of the silicon array and a plurality of
- 10 hybridization assays may be individually monitored using MALDI-TOF MS analysis.

(b) 8 x 8 (64-location) array

- A 2 X 2 cm² silicon wafer having 256 individual depressions or wells that form a 16 X 16 array of wells was purchased from a
- 15 commercial supplier [Accelerator Technology Corp., College Station, Texas]. The wells were 800 X 800 μm^2 , 120 μm deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in
- 20 iodoacetamido functionalities on the surface as described above.

- To make an array of 64 elements, a pintool was used following the procedures described above. The pintool was dipped into 16 wells of a 384 well DNA source plate containing solutions of Oligomers 1-3, moved to the silicon wafer, and the sample spotted by surface contact. Next,
- 25 the tool was dipped in washing solution, then dipped into the same 16 wells of the source plate, and spotted onto the target 2.25mm offset from the initial set of 16 spots; the entire cycle was repeated to make a 2x2 array from each pin to produce an 8x8 array of spots (2x2 elements/pin X 16 pins = 64 total elements spotted).

-223-

Oligomers 1-3 immobilized to the 64 locations were hybridized to complementary oligonucleotides and analyzed by MALDI-TOF MS analysis. As observed for the 16-location array, specific hybridization of the complementary oligonucleotide to each of the immobilized thiol-
5 containing oligomers was observed in each of the locations of the DNA array.

EXAMPLE 30

Extension of hybridized DNA primers bound to DNA templates immobilized on a silicon wafer

10 The SIAB-derivatized silicon wafers can also be employed for primer extension reactions of the immobilized DNA template using the procedures essentially described in EXAMPLE 7.

A 27-mer oligonucleotide [SEQ ID NO: 125] containing a 3'-free thiol group was coupled to a SIAB-derivatized silicon wafer as described
15 above, for example, in EXAMPLE 28. A 12-mer oligonucleotide primer [SEQ ID NO: 126] was hybridized to the immobilized oligonucleotide and the primer was extended using a commercially available kit [e.g., Sequenase or ThermoSequenase, U.S. Biochemical Corp]. The addition of Sequenase DNA polymerase or ThermoSequenase DNA polymerase in
20 the presence of three deoxyribonucleoside triphosphates (dNTPs; dATP, dGTP, dCTP) and dideoxyribonucleoside thymidine triphosphate (ddTTP) in buffer according to the instructions provided by the manufacturer resulted in a 3-base extension of the 12-mer primer while still bound to the silicon wafer. The wafer was then analyzed by MALDI-TOF mass
25 spectrometry as described above. The mass spectrum results clearly distinguish the 15-mer [SEQ ID NO: 127] from the original unextended 12-mer thus indicating that specific extension can be performed on the surface of a silicon wafer and detected using MALDI-TOF MS analysis.

-224-

EXAMPLE 31**Effect of linker length on polymerase extension of hybridized DNA primers bound to DNA templates immobilized on a silicon wafer**

The effect of the distance between the SIAB-conjugated silicon surface and the duplex DNA formed by hybridization of the target DNA to the immobilized oligomer template was investigated, as well as choice of enzyme.

Two SIAB-derivatized silicon wafers were conjugated to the 3'-end of two free thiol-containing oligonucleotides of identical DNA sequence except for a 3-base poly dT spacer sequence incorporated at the 3'-end:

10 CTGATGCGTC GGATCATCTT TTTT SEQ ID No. 122
CTGATGCGTC GGATCATCTT TTTTTT SEQ ID No. 125.

These oligonucleotides were synthesized and each was separately immobilized to the surface of a silicon wafer through the SIAB cross-linker [e.g., see EXAMPLE 28]. Each wafer was incubated with a 12-mer oligonucleotide:

15

AAAAAAGATG AT SEQ ID No. 126
GATGATCCGA CG SEQ ID No. 128
GATCCGACGC AT SEQ ID No. 129,

20 which is complementary to portions of the nucleotide sequences common to both of the oligonucleotides, by denaturing at 75 °C and slow cooling the silicon wafer. The wafers were then analyzed by MALDI-TOF mass spectrometry as described above.

As described in EXAMPLE 30 above, a 3-base specific extension of the bound 12-mer oligonucleotide was observed using the oligomer primer where there is a 9-base spacer between the duplex and the surface [SEQ ID NO: 125]. Similar results were observed when the DNA spacer lengths between the SIAB moiety and the DNA duplex were 0, 3, 6 and 12. In addition, the extension reaction may be performed using a

25

-225-

variety of DNA polymerases, such as Sequenase and Thermo Sequenase (US Biochemical). Thus, the SIAB linker may be directly coupled to the DNA template or may include a linker sequence without effecting primer extension of the hybridized DNA.

5

EXAMPLE 32**Spectrochip mutant detection in ApoE gene**

This example describes the hybridization of an immobilized template, primer extension and mass spectrometry for detection of the wildtype and mutant Apolipoprotein E gene for diagnostic purposes.

- 10 This example demonstrates that immobilized DNA molecules containing a specific sequence can be detected and distinguished using primer extension of unlabeled allele specific primers and analysis of the extension products using mass spectrometry.

A 50 base synthetic DNA template complementary to the coding

- 15 sequence of allele 3 of the wildtype apolipoprotein E gene:

5'- GCCTGGTACACTGCCAGGCGCTTCTGCAGGTCATCGGCATCGCGGAGGAG -3'
[SEQ ID NO: 280]

or complement to the mutant apolipoprotein E gene carrying a G → A transition at codon 158:

- 20 5'-GCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCGGCATCGCGGAGGAG-3'
[SEQ ID NO: 281]

containing a 3'-free thiol group was coupled to separate SIAB-derivatized silicon wafers as described in Example 28.

A 21-mer oligonucleotide primer:

- 25 5'-GAT GCC GAT GAC CTG CAG AAG-3' [SEQ ID NO: 282] was hybridized to each of the immobilized templates and the primer was extended using a commercially available kit [e.g., Sequenase or Thermosequenase, U.S. Biochemical Corp]. The addition of Sequenase DNA polymerase or Thermosequenase DNA polymerase in the presence
30 of three deoxyribonucleoside triphosphates (dNTPs; dATP, dGTP, dTTP)

-226-

and dideoxyribonucleoside cytosine triphosphate (ddCTP) in buffer according to the instructions provided by the manufacturer resulted in a single base extension of the 21-mer primer bound to the immobilized template encoding the wildtype apolipoprotein E gene and a three base extension of the 21-mer primer bound to the immobilized template encoding the mutant form of apolipoprotein E gene.

The wafers were analyzed by mass spectrometry as described herein. The wildtype apolipoprotein E sequence results in a mass spectrum that distinguishes the primer with a single base extension (22-mer) with a mass to charge ratio of 6771.17 Da (the theoretical mass to charge ratio is 6753.5 Da) from the original 21-mer primer with a mass to charge ratio of 6499.64 Da. The mutant apolipoprotein E sequence results in a mass spectrum that distinguishes the primer with a three base extension (24-mer) with a mass to charge ratio of 7386.9 (the theoretical mass charge is 7386.9) from the original 21-mer primer with a mass to charge ratio of 6499.64 Da.

EXAMPLE 33

Detection of Double-Stranded Nucleic Acid Molecules via Strand Displacement and Hybridization to an Immobilized Complementary Nucleic Acid

This example describes immobilization of a 24-mer primer and the specific hybridization of one strand of a duplex DNA molecule, thereby permitting amplification of a selected target molecule in solution phase and permitting detection of the double stranded molecule. This method is useful for detecting single base changes, and, particularly for screening genomic libraries of double-stranded fragments.

A 24-mer DNA primer CTGATGCGTC GGATCATCTT TTTT SEQ ID No. 122, containing a 3'-free thiol group was coupled to a SIAB-derivatized silicon wafer as described in Example 29.

An 18-mer synthetic oligonucleotide:

-227-

5'-CTGATGCGTCGGATCATC-3' [SEQ ID NO: 286] was premixed with a 12-mer 5'-GATGATCCGACG-3' [SEQ ID NO: 285] that has a sequence that is complementary to 12 base portion of the 18-mer oligonucleotide. The oligonucleotide mix was heated to 75°C and
5 cooled slowly to room temperature to facilitate the formation of a duplex molecule:

5' -CTGATGCGTCGGATCATC-3' [SEQ ID NO. 286]
3' - GCAGCCTAGTAG-5' [SEQ ID NO: 287].

The specific hybridization of the 12-mer strand of the duplex
10 molecule to the immobilized 24-mer primer was carried out by mixing 1μM of the duplex molecule using the hybridization conditions described in Example 30.

The wafers were analyzed by mass spectrometry as described above. Specific hybridization was detected in a mass spectrum of the
15 12-mer with a mass to charge ratio of 3682.78 Da.

EXAMPLE 34

1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

A. 2-Nitro-5-(3-hydroxypropoxy)benzaldehyde

20 3-Bromo-1-propanol (3.34 g, 24 mmol) was refluxed in 80 ml of anhydrous acetonitrile with 5-hydroxy-2-nitrobenzaldehyde (3.34 g, 20 mmol), K₂CO₃ (3.5 g), and KI (100 mg) overnight (15 h). The reaction mixture was cooled to room temperature and 150 ml of methylene chloride was added. The mixture was filtered and the solid residue was
25 washed with methylene chloride. The combined organic solution was evaporated to dryness and redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated NaCl solution and dried over sodium sulfate. 4.31 g (96%) of desired product was obtained after removal of the solvent in vacuo.
30 R_f = 0.33 (dichloromethane/methanol, 95/5).

-228-

UV (methanol) maximum: 313, 240 (shoulder), 215 nm; minimum: 266 nm.

^1H NMR (DMSO-d_6) δ 10.28 (s, 1H), 8.17 (d, 1H), 7.35 (d, 1H), 7.22 (s, 1H), 4.22 (t, 2H), 3.54 (t, 2H), 1.90 (m, 2H).

5 ^{13}C NMR (DMSO-d_6) δ 189.9, 153.0, 141.6, 134.3, 127.3, 118.4, 114.0, 66.2, 56.9, 31.7.

B. 2-Nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde

2-Nitro-5-(3-hydroxypropoxy)benzaldehyde (1 g, 4.44 mmol) was dissolved in 50 ml anhydrous acetonitrile. To this solution, it was added
10 1 ml of triethylamine, 200 mg of imidazole, and 0.8 g (5.3 mmol) of tBDMSCl. The mixture was stirred at room temperature for 4 h. Methanol (1 ml) was added to stop the reaction. The solvent was removed in vacuo and the solid residue was redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated
15 sodium bicarbonate solution and then water. The organic phase was dried over sodium sulfate and the solvent was removed in vacuo. The crude mixture was subjected to a quick silica gel column with methylene chloride to yield 1.44 g (96%) of 2-nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde.

20 R_f = 0.67 (hexane/ethyl acetate, 5/1).

UV (methanol), maximum: 317, 243, 215 nm; minimum: 235, 267 nm.

^1H NMR (DMSO-d_6) δ 10.28 (s, 1H), 8.14 (d, 1H), 7.32 (d, 1H), 7.20 (s, 1H), 4.20 (t, 2H), 3.75 (t, 2H), 1.90 (m, 2H), 0.85 (s, 9H), 0.02 (s, 6H).

^{13}C NMR (DMSO-d_6) δ 189.6, 162.7, 141.5, 134.0, 127.1, 118.2,
25 113.8, 65.4, 58.5, 31.2, 25.5, -3.1, -5.7.

C. 1-(2-Nitro-5-(3-O-t-butyldimethylsilylpropoxy)phenyl)ethanol

High vacuum dried 2-nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde (1.02 g, 3 mmol) was dissolved 50 ml of anhydrous methylene chloride. 2 M Trimethylaluminum in

-229-

toluene (3 ml) was added dropwise within 10 min and kept the reaction mixture at room temperature. It was stirred further for 10 min and the mixture was poured into 10 ml ice cooled water. The emulsion was separated from water phase and dried over 100 g of sodium sulfate to remove the remaining water. The solvent was removed in vacuo and the mixture was applied to a silica gel column with gradient methanol in methylene chloride. 0.94 g (86%) of desired product was isolated.

R_f = 0.375 (hexane/ethyl acetate, 5/1).

UV (methanol), maximum: 306, 233, 206 nm; minimum: 255, 220 nm.

¹H NMR (DMSO- d_6) δ 8.00 (d, 1H), 7.36 (s, 1H), 7.00 (d, 1H), 5.49 (b, OH), 5.31 (q, 1H), 4.19 (m, 2H), 3.77 (t, 2H), 1.95 (m, 2H), 1.37 (d, 3H), 0.86 (s, 9H), 0.04 (s, 6H).

¹³C NMR (DMSO- d_6) δ 162.6, 146.2, 139.6, 126.9, 112.9, 112.5, 64.8, 63.9, 58.7, 31.5, 25.6, 24.9, -3.4, -5.8.

15 D. 1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol

1-(2-Nitro-5-(3-O-t-butyldimethylsilylpropoxy)phenyl)ethanol (0.89 g, 2.5 mmol) was dissolved in 30 ml of THF and 0.5 mmol of $n\text{Bu}_4\text{NF}$ was added under stirring. The mixture was stirred at room temperature for 5 h and the solvent was removed in vacuo. The remaining residue was applied to a silica gel column with gradient methanol in methylene chloride. 1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol (0.6 g (99%)) was obtained.

R_f = 0.17 (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 304, 232, 210 nm; minimum: 255, 219 nm.

¹H NMR (DMSO- d_6) δ 8.00 (d, 1H), 7.33 (s, 1H), 7.00 (d, 1H), 5.50 (d, OH), 5.28 (t, OH), 4.59 (t, 1H), 4.17 (t, 2H), 3.57 (m, 2H), 1.89 (m, 2H), 1.36 (d, 2H).

¹³C NMR (DMOS- d_6) δ 162.8, 146.3, 139.7, 127.1, 113.1, 112.6, 65.5, 64.0, 57.0, 31.8, 25.0.

-230-

E. 1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol

1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol (0.482 g, 2 mmol) was co-evaporated with anhydrous pyridine twice and dissolved in 20 ml anhydrous pyridine. The solution was cooled in ice-water bath and 750
5 mg (2.2 mmol) of DMTCl was added. The reaction mixture was stirred at room temperature overnight and 0.5 ml methanol was added to stop the reaction. The solvent was removed in vacuo and the residue was co-evaporated with toluene twice to remove trace of pyridine. The final
10 residue was applied to a silica gel column with gradient methanol in methylene chloride containing drops of triethylamine to yield 0.96 g (89%) of the desired product 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol.

R_f = 0.50 (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 350 (shoulder), 305, 283, 276 (shoulder),
15 233, 208 nm; minimum: 290, 258, 220 nm.
 ^1H NMR (DMSO- d_6) δ 8.00 (d, 1H), 6.82-7.42 (ArH), 5.52 (d, OH), 5.32 (m, 1H), 4.23 (t, 2H), 3.71 (s, 6H), 3.17 (t, 2H), 2.00 (m, 2H), 1.37 (d, 3H).

^{13}C NMR (DMOS- d_6) δ 162.5, 157.9, 157.7, 146.1, 144.9, 140.1,
20 139.7, 135.7, 129.5, 128.8, 127.6, 127.5, 127.3, 126.9, 126.4, 113.0, 112.8, 112.6, 85.2, 65.3, 63.9, 59.0, 54.8, 28.9, 24.9.

F. 1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol (400
25 mg, 0.74 mmol) was dried under high vacuum and was dissolved in 20 ml of anhydrous methylene chloride. To this solution, it was added 0.5 ml N,N-diisopropylethylamine and 0.3 ml (1.34 mmol) of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite. The reaction mixture was stirred at room temperature for 30 min and 0.5 ml of methanol was added to

-231-

stop the reaction. The mixture was washed with saturated sodium bicarbonate solution and was dried over sodium sulfate. The solvent was removed in vacuo and a quick silica gel column with 1% methanol in methylene chloride containing drops of triethylamine yield 510 mg (93%)

5 the desired phosphoramidite.

$R_f = 0.87$ (dichloromethane/methanol, 99/1).

EXAMPLE 35

1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

10 A. 4-(3-Hydroxypropoxy)-3-methoxyacetophenone

3-Bromo-1-propanol (53 ml, 33 mmol) was refluxed in 100 ml of anhydrous acetonitrile with 4-hydroxy-3-methoxyacetophenone (5 g, 30 mmol), K_2CO_3 (5 g), and KI (300 mg) overnight (15 h).

Methylenechloride (150 ml) was added to the reaction mixture after
15 cooling to room temperature. The mixture was filtered and the solid residue was washed with methylene chloride. The combined organic solution was evaporated to dryness and redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated NaCl solution and dried over sodium sulfate. 6.5 g (96.4%) of desired product was
20 obtained after removal of the solvent in vacuo.

$R_f = 0.41$ (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 304, 273, 227, 210 nm; minimum: 291, 244, 214 nm.

1H NMR (DMSO- d_6) δ 7.64 (d, 1H), 7.46 (s, 1H), 7.04 (d, 1H), 4.58 (b,
25 OH), 4.12 (t, 2H), 3.80 (s, 3H), 3.56 (t, 2H), 2.54 (s, 3H), 1.88 (m, 2H).

^{13}C NMR (DMSO- d_6) δ 196.3, 152.5, 148.6, 129.7, 123.1, 111.5, 110.3, 65.4, 57.2, 55.5, 31.9, 26.3.

B. 4-(3-Acetoxypropoxy)-3-methoxyacetophenone

-232-

4-(3-Hydroxypropoxy)-3-methoxyacetophenone (3.5 g, 15.6 mmol) was dried and dissolved in 80 ml anhydrous acetonitrile. This mixture, 6 ml of triethylamine and 6 ml of acetic anhydride were added. After 4 h, 6 ml methanol was added and the solvent was removed in vacuo. The

- 5 residue was dissolved in 100 ml dichloromethane and the solution was washed with dilute sodium bicarbonate solution, then water. The organic phase was dried over sodium sulfate and the solvent was removed. The solid residue was applied to a silica gel column with methylene chloride to yield 4.1 g of 4-(3-acetoxypropoxy)-3-methoxyacetophenone (98.6%).

$R_f = 0.22$ (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 303, 273, 227, 210 nm; minimum: 290, 243, 214 nm.

- ^1H NMR (DMSO- d_6) δ 7.62 (d, 1H), 7.45 (s, 1H), 7.08 (d, 1H), 4.12 (m, 4H, 3.82 (s, 3H), 2.54 (s, 3H), 2.04 (m, 2H), 2.00 (s, 3H).

^{13}C NMR (DMSO- d_6) δ 196.3, 170.4, 152.2, 148.6, 130.0, 123.0, 111.8, 110.4, 65.2, 60.8, 55.5, 27.9, 26.3, 20.7.

C. 4-(3-Acetoxypropoxy)-3-methoxy-6-nitroacetophenone

- 4-(3-Acetoxypropoxy)-3-methoxyacetophenone (3.99 g, 15 mmol) was added portionwise to 15 ml of 70% HNO_3 in water bath and keep the reaction temperature at the room temperature. The reaction mixture was stirred at room temperature for 30 min and 30 g of crushed ice was added. This mixture was extracted with 100 ml of dichloromethane and the organic phase was washed with saturated sodium bicarbonate solution. The solution was dried over sodium sulfate and the solvent was removed in vacuo. The crude mixture was applied to a silica gel column with gradient methanol in methylene chloride to yield 3.8 g (81.5%) of desired product 4-(3-acetoxypropoxy)-3-methoxy-6-

-233-

nitroacetophenone and 0.38 g (8%) of ipso-substituted product 5-(3-acetoxypropoxy)-4-methoxy-1,2-dinitrobenzene.

Side ipso-substituted product 5-(3-acetoxypropoxy)-4-methoxy-1,2-dinitrobenzene:

5 $R_f = 0.47$ (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 334, 330, 270, 240, 212 nm; minimum: 310, 282, 263, 223 nm.

^1H NMR (CDCl_3) δ 7.36 (s, 1H), 7.34 (s, 1H), 4.28 (t, 2H), 4.18 (t, 2H), 4.02 (s, 3H), 2.20 (m, 2H), 2.08 (s, 3H).

10 ^{13}C NMR (CDCl_3) δ 170.9, 152.2, 151.1, 117.6, 111.2, 107.9, 107.1, 66.7, 60.6, 56.9, 28.2, 20.9.

Desired product 4-(3-acetoxypropoxy)-3-methoxy-6-nitroacetophenone:

$R_f = 0.29$ (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 344, 300, 246, 213 nm; minimum: 320,

15 270, 227 nm.

^1H NMR (CDCl_3) δ 7.62 (s, 1H), 6.74 (s, 1H), 4.28 (t, 2H), 4.20 (t, 2H), 3.96 (s, 3H), 2.48 (s, 3H), 2.20 (m, 2H), 2.08 (s, 3H).

^{13}C NMR (CDCl_3) δ 200.0, 171.0, 154.3, 148.8, 138.3, 133.0, 108.8, 108.0, 66.1, 60.8, 56.6, 30.4, 28.2, 20.9.

20 **D. 1-(4-(3-Hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol**

4-(3-Acetoxypropoxy)-3-methoxy-6-nitroacetophenone (3.73 g, 12 mmol) was added 150 ml ethanol and 6.5 g of K_2CO_3 . The mixture was stirred at room temperature for 4h and TLC with 5% methanol in dichloromethane indicated the completion of the reaction. To this same

25 reaction mixture, it was added 3.5 g of NaBH_4 and the mixture was stirred at room temperature for 2h. Acetone (10 ml) was added to react with the remaining NaBH_4 . The solvent was removed in vacuo and the residue was uptaken into 50 g of silica gel. The silica gel mixture was applied on the top of a silica gel column with 5% methanol in methylene

-234-

chloride to yield 3.15 g (97%) of desired product 1-(4-(3-hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol.

Intermediate product 4-(3-hydroxypropoxy)-3-methoxy-6-nitroacetophenone after deprotection:

- 5 $R_f = 0.60$ (dichloromethane/methanol, 95/5).

Final product 1-(4-(3-hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol:

$R_f = 0.50$ (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 344, 300, 243, 219 nm; minimum: 317, 264, 233 nm.

- 10 ^1H NMR (DMSO- d_6) δ 7.54 (s, 1H), 7.36 (s, 1H), 5.47 (d, OH), 5.27 (m, 1H), 4.55 (t, OH), 4.05 (t, 2H), 3.90 (s, 3H), 3.55 (q, 2H), 1.88 (m, 2H), 1.37 (d, 3H).

^{13}C NMR (DMSO- d_6) δ 153.4, 146.4, 138.8, 137.9, 109.0, 108.1, 68.5, 65.9, 57.2, 56.0, 31.9, 29.6.

- 15 E. 1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol

1-(4-(3-Hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol (0.325 g, 1.2 mmol) was co-evaporated with anhydrous pyridine twice and dissolved in 15 ml anhydrous pyridine. The solution was cooled in ice-

- 20 water bath and 450 mg (1.33 mmol) of DMTCl was added. The reaction mixture was stirred at room temperature overnight and 0.5 ml methanol was added to stop the reaction. The solvent was removed in vacuo and the residue was co-evaporated with toluene twice to remove trace of pyridine. The final residue was applied to a silica gel column with
- 25 gradient methanol in methylene chloride containing drops of triethylamine to yield 605 mg (88%) of desired product 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol.

$R_f = 0.50$ (dichloromethane/methanol, 95/5).

-235-

UV (methanol), maximum: 354, 302, 282, 274, 233, 209 nm; minimum: 322, 292, 263, 222 nm.

^1H NMR ($\text{DMSO}-d_6$) δ 7.54 (s, 1H), 6.8-7.4 (ArH), 5.48 (d, OH), 5.27 (m, 1H), 4.16 (t, 2H), 3.85 (s, 3H), 3.72 (s, 6H), 3.15 (t, 2H), 1.98 (t, 2H),

5 1.37 (d, 3H).

^{13}C NMR ($\text{DMSO}-d_6$) δ 157.8, 153.3, 146.1, 144.9, 138.7, 137.8, 135.7, 129.4, 128.7, 127.5, 127.4, 126.3, 112.9, 112.6, 108.9, 108.2, 85.1, 65.7, 63.7, 59.2, 55.8, 54.8, 29.0, 25.0.

10 F. 1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol (200 mg, 3.5 mmol) was dried under high vacuum and was dissolved in 15 ml of anhydrous methylene chloride. To this
15 solution, it was added 0.5 ml N,N-diisopropylethylamine and 0.2 ml (0.89 mmol) of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite. The reaction mixture was stirred at room temperature for 30 min and 0.5 ml of methanol was added to stop the reaction. The mixture was washed with saturated sodium bicarbonate solution and was dried over sodium
20 sulfate. The solvent was removed in vacuo and a quick silica gel column with 1% methanol in methylene chloride containing drops of triethylamine yield 247 mg (91.3%) the desired phosphoramidite 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

25 $R_f = 0.87$ (dichloromethane/methanol, 99/1).

-236-

EXAMPLE 36

Oligonucleotide synthesis

- The oligonucleotide conjugates containing photocleavable linker were prepared by solid phase nucleic acid synthesis (see: Sinha *et al.* Tetrahedron Lett. **1983**, 24, 5843-5846; Sinha *et al.* Nucleic Acids Res. **1984**, 12, 4539-4557; Beaucage *et al.* Tetrahedron **1993**, 49, 6123-6194; and Matteucci *et al.* J. Am. Chem. Soc. **1981**, 103, 3185-3191) under standard conditions. In addition a longer coupling time period was employed for the incorporation of photocleavable unit and the 5' terminal amino group. The coupling efficiency was detected by measuring the absorbance of released DMT cation and the results indicated a comparable coupling efficiency of phosphoramidite 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane or 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane with those of common nucleoside phosphoramidites. Deprotection of the base protection and release of the conjugates from the solid support was carried out with concentrated ammonium at 55 °C overnight. Deprotection of the base protection of other conjugates was done by fast deprotection with AMA reagents. Purification of the MMT-on conjugates was done by HPLC (trityl-on) using 0.1 M triethylammonium acetate, pH 7.0 and a gradient of acetonitrile (5% to 25% in 20 minutes). The collected MMT or DMT protected conjugate was reduced in volume, detritylated with 80% aqueous acetic acid (40 min, 0 °C), desalted, stored at -20°C.

-237-

EXAMPLE 37**Photolysis study**

- In a typical case, 2 nmol of oligonucleotide conjugate containing photocleavable linker in 200 μ l distilled water was irradiated with a long
- 5 wavelength UV lamp (Blak Ray XX-15 UV lamp, Ultraviolet products, San Gabriel, CA) at a distance of 10 cm (emission peak 365 nm, lamp intensity = 1.1 mW/cm² at a distance of 31 cm). The resulting mixture was analyzed by HPLC (trityl-off) using 0.1 M triethylammonium acetate, pH 7.0 and a gradient of acetonitrile. Analysis showed that the
- 10 conjugate was cleaved from the linker within minutes upon UV irradiation.

Equivalents

- Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the
- 15 specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

-238-

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT:

(A) NAME: SEQUENOM, INC.
(B) STREET: 11555 Sorrento Valley Road
(C) CITY: San Diego
(D) STATE: California
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 92121

(i) INVENTOR/APPLICANT:

(A) NAME: Hubert Köster
(B) STREET: 836 Via Mallorca Drive
(C) CITY: La Jolla
(D) STATE: California
(D) COUNTRY: USA
(E) POSTAL CODE (ZIP): 92037

(i) INVENTOR/APPLICANT:

(A) NAME: Kai Tang
(B) STREET: 8521 Summerdale Rd #241
(C) CITY: San Diego
(D) STATE: California
(D) COUNTRY: USA
(E) POSTAL CODE (ZIP): 92126

(i) INVENTOR/APPLICANT:

(A) NAME: Long-Jing Fu
(B) STREET: 10615 Dabney Dr. #21
(C) CITY: San Diego
(D) STATE: California
(D) COUNTRY: USA
(E) POSTAL CODE (ZIP): 92126

(i) INVENTOR/APPLICANT:

(A) NAME: Carsten W. Siegert
(B) STREET: Geilstr. 42
(C) CITY: 22303 Hamburg
(D) STATE:
(D) COUNTRY: Germany
(E) POSTAL CODE (ZIP):

(i) INVENTOR/APPLICANT:

(A) NAME: Daniel P. Little
(B) STREET: 393 Glendale Lake Rd.
(C) CITY: Patton
(D) STATE: Pennsylvania
(D) COUNTRY: USA
(E) POSTAL CODE (ZIP): 18668

(i) INVENTOR/APPLICANT:

(A) NAME: G. Scott Higgins
(B) STREET: Haselweg 1
(C) CITY: 22880 Weidel
(D) STATE:
(D) COUNTRY: Germany
(E) POSTAL CODE (ZIP):

-239-

- (i) INVENTOR/APPLICANT:
 (A) NAME: Andreas Braun
 (B) STREET: 13232 Benchley Road
 (C) CITY: San Diego
 (D) STATE: California
 (E) COUNTRY: USA
 (E) POSTAL CODE (ZIP): 92130
- (i) INVENTOR/APPLICANT:
 (A) NAME: Brigitte Damhoffer-Demar
 (B) STREET: 3899 Haines St. #8-308
 (C) CITY: San Diego
 (D) STATE: California
 (E) COUNTRY: USA
 (E) POSTAL CODE (ZIP): 92109
- (i) INVENTOR/APPLICANT:
 (A) NAME: Christian Jurinke
 (B) STREET: Grope Hall 68
 (C) CITY: 22115 Hamburg
 (D) STATE:
 (D) COUNTRY: Germany
 (E) POSTAL CODE (ZIP):
- (i) INVENTOR/APPLICANT:
 (A) NAME: Dirk Van den Boom
 (B) STREET: Forsthausstr. 8
 (C) CITY: 633303 Preiech
 (D) STATE:
 (D) COUNTRY: Germany
 (E) POSTAL CODE (ZIP):
- (i) INVENTOR/APPLICANT:
 (A) NAME: Goubing Xiang
 (B) STREET: 11381 Zapata Ave.
 (C) CITY: San Diego
 (D) STATE: California
 (E) COUNTRY: USA
 (E) POSTAL CODE (ZIP): 92126
- (ii) TITLE OF THE INVENTION: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY
- (iii) NUMBER OF SEQUENCES: 320
- (iv) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: Brown, Martin, Haller & McClain
 (B) STREET: 1660 Union Street
 (C) CITY: San Diego
 (D) STATE: CA
 (E) COUNTRY: USA
 (F) ZIP: 92101-2926
- (v) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: Diskette
 (B) COMPUTER: IBM Compatible
 (C) OPERATING SYSTEM: DOS
 (D) SOFTWARE: FastSEQ Version 1.5
- (vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER:

-240-

- (B) FILING DATE: 06-NOV-1997
- (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: 10/08/97
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/933,792
 - (B) FILING DATE: 09/19/97
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/787,639
 - (B) FILING DATE: 01/23/97
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/786,988
 - (B) FILING DATE: 01/23/97
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/746,055
 - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/746,036
 - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/744,590
 - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/744,481
 - (B) FILING DATE: 11/06/96
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Seidman, Stephanie L
 - (B) REGISTRATION NUMBER: 33,779
 - (C) REFERENCE/DOCKET NUMBER: 7352-2004PC
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 619-238-0999
 - (B) TELEFAX: 619-238-0062
 - (C) TELEX:
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTISENSE: NO
 - (v) FRAGMENT TYPE:
 - (vi) ORIGINAL SOURCE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

-241-

GCAAGTGAAT CCTGAGCGTG

20

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGTGAAGGG TTCATATGC

19

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCTATATTC ATCATAGGAA ACACCACA

28

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTATCTATAT TCATCATAGG AACACCATT

30

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid

-242-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCTTTGGGGC ATGGACATTG ACCCGTATAA

30

- (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTGACTACTA ATTCCCTGGA TGCTGGGTCT

30

- (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTGCCTGAGT GCAGTATGGT

20

- (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-243-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGCTCTATAT CGGGAAGCCT

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTGTGCCACG CGGTTGGGAA TGTA

24

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCAACGACT GTTTGCCCCG CAGTTG

26

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

-244-

TACATTCCCA ACCGCGTGGC ACAAC

25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AACTGGCGGG CAAACAGTCG TTGCT

25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCAAGTGAAT CCTGAGCGTG

20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGTGAAGGG CGTG

14

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs

-245-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTATATTCAT CATAGGAAAC ACCA

24

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTCACCCTCG ACCTCCAG

18

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGTAAAACG ACGGCCAGT

19

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-246-

(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTTCCACCGC GATGTTGA

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAGGAAACAG CTATGAC

17

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAAACGAC GGCCAGT

17

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTCACCCTCG ACCTGCAGC

19

-247-

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GTTGTAAAAC GAGGGCCAGT

20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCTGGCCTGG TGCAGGGCCT ATTGTAGTTG TGACGTACA

39

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TGTACGTCAC AACT

14

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-248-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAGATCTGAC CAGGGATTTCG GTTAGCGTGA CTGCTGCTGC TGCTGCTGCT GCTGGATGAT 60
CCGACGCATC AGATCTGG

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATGATCCGA CGCATCACAG CTC

23

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-249-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TCGGTTCCAA GAGCTGTGAT GCGTCGGATC ATC

33

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GATGATCCGA CGCATCACAG CTC

23

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCGGTTCCAA GAGCT

15

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

-250-

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCGGTTCCAA GAGCT

15

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CATTGCTTC TGACACAACT G

21

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTTCTCTGTC TCCACATGC

19

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-251-

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TGCACCTGAC TC

12

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGCTTACTTA ACCCAGTGTG

20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CACACTATGT AATACTATGC

20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

-252-

GAAAATATCT GACAAACTCA TC

22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATGGACACC AAATTAAGTT C

21

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TGAGACTCTG TCTC

14

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TTCCCCAAAT CCCTG

15

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs

-253-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGCACGGCTG TCCAAGGAG

19

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCGGACATGG AGGACGTG

18

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-254-

(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CCCTTACCCT TACCCTTACC CTAA

24

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AATCCGTGCA GCAGAGTT

18

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TGTCAGAGCT GGACAAGTGT

20

-255-

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATATTGTCT TCCCGGTAGC

20

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTCGGACCAG GTGTACCGCC

20

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCTGTACTGG AAGGCGATCT C

21

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-256-

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CATGAGGCAG AGCATACGCA

20

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GACAGCAGCA CCGAGACGAT

20

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CGGCTGCGAT CACCGTGCGG

20

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

-257-

GATCCACTGT GCGACGAGC

19

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCGGCTGCGA TCACCGTGC

19

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TGCACCTGAC TC

12

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CTGTGGTCGT GC

12

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-258-

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GAGTCAGGTG CGCCATGCCT CAAACAGACA CCATGGCGC

39

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TCTCTGTCTC CACATGCCCA G

21

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ACCTAGCGTT CAGTTCGACT GAGATAATAC GACTCACTAT AGCAGCTCTC ATTTTCCATA 60
C 61

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:

-259-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AACTAAGCCA TGTGCACAAC A

21

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

UCCGGUCUGA UGAGUCCGUG AGGAC

25

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GUCACUACAG GUGAGCUCCA

20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CCAUGCGAGA GUAAGUAGUA

20

(2) INFORMATION FOR SEQ ID NO:65:

-260-

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: RNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

AGGCCUGCGG CAAGACGGAA AGACCAUGGU CCCUNAUCUG CCGCAGGAUC

50

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CATTTGCTTC TGACACAACT

20

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

TCTCTGTCTC CACATGCCCA G

21

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA

-261-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTCGTCCCAT GGTGCACCTG ACTC

24

(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGCTGTGGTG AGGCCCTGGG CA

22

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GACGACGACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

-262-

ACAGCGGACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TGGAGTCAGG TAGCAGTC

18

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CAGCTCTCAT TTTCCATAC

19

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

AGCCCCAAGA TGACTATC

18

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid

-263-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CGAGGAGCTC AAGGCCAGAA T

21

- (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CAGGGGCAGC TCAGCTCTC

19

- (2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGCACGGCTG TCCAAGGA

18

- (2) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-264-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CTTACTTGAA TTCCAAGAGC

20

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGGCTGACTT GCATGGACCG GA

22

(2) INFORMATION FOR SEQ ID NO:81

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

AGCCAGGACA AG

12

-265-

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

ACAGCAGGAA CAGCA

15

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GCGGACATGG AGGACGTG

18

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-266-

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

GTGCCCTGCA GCTTCACTGA AGAC

24

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AGCCAGGACA AG

12

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

AGCCAGGACA AGTC

14

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

-267-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

AGCCAGGACA AGA

13

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ACAGCACCAA CAGCA

15

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

ACAGCAGGAA CAGCATC

17

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

ACAGCAGGAA CAGCAG

16

-268-

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GCGGACATGG AGGACGTG

18

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

GCGGACATGG AGGACGTGGC

20

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GCGGACATGG AGGACGTGC

19

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-269-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

GATGCCGATG ACCTGCAGAA GC

22

(2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GATGCCGATG ACCTGCAGAA GTG

23

(2) INFORMATION FOR SEQ ID NO:98:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-270-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GTGCCCTGCA GCTTCACTGA AGAC

24

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GTGCCCTGCA GCTTCACTGA AGACTG

26

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GTGCCCTGCA GCTTCACTGA AGACC

25

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TATCTGTTCA CTTGTGCCC

19

-271-

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CAGAGGCCTG GGGACCCTG

19

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ACGACAGGGC TGGTTGCC

18

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

ACTGACAACC ACCCTTAAC

19

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid

-272-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

CTGCTTGCCA CAGGTCTC

18

(2) INFORMATION FOR SEQ ID NO:106:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

CACAGCAGGC CAGTGTGC

18

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

GGACCTGATT TCCTTACTG

19

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA

-273-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

TGAATCTGAG GCATAACTG

19

(2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TTGCGTACAC ACTGGCCGTC GTTTTACAAC GTCGTGACTG GGAAAACCCT

50

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

GTAAAACGAC GGCCAGTGTG TACGCAA

27

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-274-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

TACTGGAAGG CGATCTCAGC AATCAGC

27

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

GGCACGGCTG TCCAAGGAG

19

(189) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

GUCACUACAG GUGAGCUCCA

20

(2) INFORMATION FOR SEQ ID NO:115:

-275-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GAATTCGAGC TCGGTACCCG G

21

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

CCGGGTACCG AGCTCGAATT C

21

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CCTCTTGGA ACTGTGTAGT ATT

23

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-276-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

AGGCTGTCTC TCTCCCTCTC TCATACACAC ACACACACAC ACACACACAC ACACACACAC 60
ACACACACAC TCACACTCAC CCACANNNA AATACTACACA GTTCCCAAGA GG 112

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

TAATACGACT CACTATAGGG CGAAGGCTGT CTCTCTCCCT CTCTCATAC 49

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

TAATACGACT CACTATAGGG CGAAGGCTGT CTCTCTCCCT CTCTCATACA CACACACACA 60
CACACACACA CACACACACA CACTCACACT CACCCACANN NAAATACTAC 120
ACAGTTCCTCA AGAGG 135

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

-277-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

AATACTACAC AG

12

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

CTGATGCGTC GGATCATCTT TTTT

24

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GATGATCCGA CGCATCAGAA TGT

23

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

GATCTAGCTG GGCCGAGCTA GGCCGTTGA

29

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

-278-

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CTGATGCGTC GGATCATCTT TTTTTTT

27

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

GATGATCCGA CG

12

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

GATGATCCGA CGCAT

15

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-279-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

AAAAAAGATG AT

12

(2) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GATCCGACGC AT

12

(2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 253 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

```
GGCACGGCTG TCCAAGGAGC TGCAGGCGGC GCAGGCCCGG CTGGGCGCGG ACATGGAGGA 60
CGTGTGCGGC CGCCTGGTGC AGTACCGCGG CGAGGTGCAG GCCATGCTCG GCCAGAGCAC 120
CGAGGAGCTG CGGGTGCGCC TCGCCTCCCA CCTGCGCAAG CTGCGTAAGC GGCTCCTCCG 180
CGATGCCGAT GACCTGCAGA AGTGCCTGGC AGTGTACCAG GCCGGGGCCC GCGAGGGCGC 240
CGAGCGCGGC CTC 253
```

(2) INFORMATION FOR SEQ ID NO:131:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 58 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-280-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GAATTACATT CCCAACCGCG TGGCACAACA ACTGGCGGGC AAACAGTCGT TGCTGATT 58

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

ACCATTAAAG AAAATATCAT CTTTGGTGTT TCCTATGATG AATATAGAAG CGTCATC 57

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

CTATATTCAT CATAGGAAAC ACCAAAGAT 29

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

CTATATTCAT CATAGGAAAC ACCAAT 26

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

-281-

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

CTATATTCAT CATAGGAAAC ACCAAAGAT

29

(2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CTATATTCAT CATAGGAAAC ACCAAAGATG ATATTTTC

38

(2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CTATATTCAT CATAGGAAAC ACCAATG ATATTTTC

35

(2) INFORMATION FOR SEQ ID NO:138:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-282-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

CTATATTCAT CATAGGAAAC ACCAAAGATA TTTTC

35

(2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

CTATATTCAT CATAGGAAAC ACCAAAGATG C

31

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 200 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

CTTCCACCGC GATGTTGATG ATTATGTGTC TGAATTTGAT GGGGGCAGGC GGCCCCCGTC
TGTTTGTCGC GGGTCTGGTG TTGATGGTGG TTTCCTGCCT TGTCACCCTC GACCTGCAGC
CCAAGCTTGG GATCCACCAC CATCACCATC ACTAATAATG CATGGGCTGC AGCCAATTGG
CACTGGCCGT CGTTTACAA

(2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 99 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-283-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GTCACCCTCG ACCTGCAGCC CAAGCTTGGG ATCCACCACC ATCACCATCA CTAATAATGC
ATGGGCTGCA GCCAATTGGC ACTGGCCGTC GTTTTACAA

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

TGTACGTCAC AACTA

15

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

TGTACGTCAC AACTAC

16

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

TGTACGTCAC AACTACA

17

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

-284-

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

TGACGTCAC AACTACAA

18

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

TGACGTCAC AACTACAAT

19

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TGACGTCAC AACTACAATA

20

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-285-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

TGTACGTCAC AACTACAATA G

21

(2) INFORMATION FOR SEQ ID NO:149:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TGTACGTCAC AACTACAATA GG

22

(2) INFORMATION FOR SEQ ID NO:150:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

TGTACGTCAC AACTACAATA GGC

23

(2) INFORMATION FOR SEQ ID NO:151:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

TGTACGTCAC AACTACAATA GGCC

24

-286-

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

TGTACGTCAC AACTACAATA GGCCC

25

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

TGTACGTCAC AACTACAATA GGCCCT

26

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTACGTCAC AACTACAATA GGCCCTG

27

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-287-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

TGTACGTCAC AACTACAATA GGCCCTGC

28

(2) INFORMATION FOR SEQ ID NO:156:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

TGTACGTCAC AACTACAATA GGCCCTGCA

29

(2) INFORMATION FOR SEQ ID NO:157:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

TGTACGTCAC AACTACAATA GGCCCTGCAC

30

(2) INFORMATION FOR SEQ ID NO:158:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

-288-

TGTACGTCAC AACTACAATA GGCCCTGCAC C

31

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TGTACGTCAC AACTACAATA GGCCCTGCAC CA

32

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAG

33

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGG

34

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid

-289-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGC

35

- (2) INFORMATION FOR SEQ ID NO:163:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCC

36

- (2) INFORMATION FOR SEQ ID NO:164:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCA

37

- (2) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-290-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCAG

38

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCAGA

39

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

CTGATGCGTC GGATCATCC

19

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

CTGATGCGTC GGATCATCCA

20

(2) INFORMATION FOR SEQ ID NO:169:

-291-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

CTGATGCGTC GGATCATCCA G

21

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

CTGATGCGTC GGATCATCCA GC

22

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

CTGATGCGTC GGATCATCCA GCA

23

(2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-292-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

CTGATGCGTC GGATCATCCA GCAG

24

(2) INFORMATION FOR SEQ ID NO:173:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

CTGATGCGTC GGATCATCCA GCAGC

25

(2) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

CTGATGCGTC GGATCATCCA GCAGCA

26

(2) INFORMATION FOR SEQ ID NO:175:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

-293-

CTGATGCGTC GGATCATCCA GCAGCAG

27

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

CTGATGCGTC GGATCATCCA GCAGCAGC

28

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

CTGATGCGTC GGATCATCCA GCAGCAGCA

29

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CTGATGCGTC GGATCATCCA GCAGCAGCAG

30

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid

-294-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

CTGATGCGTC GGATCATCCA GCAGCAGCAG C

31

- (2) INFORMATION FOR SEQ ID NO:180:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CA

32

- (2) INFORMATION FOR SEQ ID NO:181:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAG

33

- (2) INFORMATION FOR SEQ ID NO:182:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-295-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGC

34

(2) INFORMATION FOR SEQ ID NO:183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCA

35

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAG

36

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGC

37

(2) INFORMATION FOR SEQ ID NO:186:

-296-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCA

38

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAG

39

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC

40

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-297-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC A

41

(2) INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AG

42

(2) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGT

43

(2) INFORMATION FOR SEQ ID NO:192:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

-298-

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTC

44

(2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCA

45

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCAC

46

(2) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACG

47

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid

-299-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGC

48

- (2) INFORMATION FOR SEQ ID NO:197:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 49 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCT

49

- (2) INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA

50

- (2) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-300-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA A

51

(2) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA AC

52

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACC

53

(2) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCG

54

(2) INFORMATION FOR SEQ ID NO:203:

-301-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGA

55

(2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAA

56

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAAT

57

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-302-

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATC 58

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCC 59

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

-303-

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
T 61

(2) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TG 62

(2) INFORMATION FOR SEQ ID NO:211:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGG 63

(2) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGT 64

(2) INFORMATION FOR SEQ ID NO:213:

-304-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTC 65

(2) INFORMATION FOR SEQ ID NO:214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCA 66

(2) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAG 67

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-305-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAGA 68

(2) INFORMATION FOR SEQ ID NO:217:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAGAT 69

(2) INFORMATION FOR SEQ ID NO:218:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAGATC 70

(2) INFORMATION FOR SEQ ID NO:219:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 71 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-306-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAGATC T 71

(2) INFORMATION FOR SEQ ID NO:220:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60
TGGTCAGATC TT 72

(2) INFORMATION FOR SEQ ID NO:221:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

TGCACCTGAC TCC 13

(2) INFORMATION FOR SEQ ID NO:222:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

-307-

TGCACCTGAC TCCT

14

(2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

TGCACCTGAC TCCTG

15

(2) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

TGCACCTGAC TCCTGT

16

(2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

TGCACCTGAC TCCTGTG

17

(2) INFORMATION FOR SEQ ID NO:226:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid

-308-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

TGCACCTGAC TCCTGTGG

18

(2) INFORMATION FOR SEQ ID NO:227:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

TGCACCTGAC TCCTGTGGA

19

(2) INFORMATION FOR SEQ ID NO:228:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

TGCACCTGAC TCCTGTGGAG

20

(2) INFORMATION FOR SEQ ID NO:229:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-309-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

TGCACCTGAC TCCTGTGGAG A

21

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

TGCACCTGAC TCCTGTGGAG AA

22

(2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

TGCACCTGAC TCCTGTGGAG AAG

23

(2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

TGCACCTGAC TCCTGTGGAG AAGT

24

(2) INFORMATION FOR SEQ ID NO:233:

-310-

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

TGCACCTGAC TCCTGTGGAG AAGTC

25

(2) INFORMATION FOR SEQ ID NO:234:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

TGCACCTGAC TCCTGTGGAG AAGTCT

26

(2) INFORMATION FOR SEQ ID NO:235:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

TGCACCTGAC TCCTGTGGAG AAGTCTG

27

(2) INFORMATION FOR SEQ ID NO:236:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA

-311-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

TGCACCTGAC TCCTGTGGAG AAGTCTGC

28

(2) INFORMATION FOR SEQ ID NO:237:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

TGCACCTGAC TCCTGTGGAG AAGTCTGCC

29

(2) INFORMATION FOR SEQ ID NO:238:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG

30

(2) INFORMATION FOR SEQ ID NO:239:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

-312-

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG T

31

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TT

32

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA

33

(2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAC

34

(2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid

-313-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACT

35

- (2) INFORMATION FOR SEQ ID NO:244:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTG

36

- (2) INFORMATION FOR SEQ ID NO:245:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGC

37

- (2) INFORMATION FOR SEQ ID NO:246:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-314-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCC

38

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCC

39

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT

40

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT G

41

(2) INFORMATION FOR SEQ ID NO:250:

-315-

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GT

42

(2) INFORMATION FOR SEQ ID NO:251:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTG

43

(2) INFORMATION FOR SEQ ID NO:252:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGG

44

(2) INFORMATION FOR SEQ ID NO:253:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA

-316-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGG

45

(2) INFORMATION FOR SEQ ID NO:254:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGG

46

(2) INFORMATION FOR SEQ ID NO:255:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGC

47

(2) INFORMATION FOR SEQ ID NO:256:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

-317-

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCA

48

(2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCAA

49

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCAAG

50

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCAAG G

51

(2) INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid

-318-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCAAG GT 52

(2) INFORMATION FOR SEQ ID NO:261:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 209 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

CATTGCTTC TGACACAACT GTGTTCACTA GCAACCTCAA ACAGACACCA TGGTGACCT 60
GACTCCTGTG GAGAAGTCTG CCGTTACTGC CCTGTGGGGC AAGGTGAACG TGGATGAAGT 120
TGGTGGTGAG GCCCTGGGCA GGTGGTATC AAGGTTACAA GACAGGTTTA AGGAGACCAA 180
TAGAACTGG GCATGTGGAG ACAGAGAAG 209

(2) INFORMATION FOR SEQ ID NO:262:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 88 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

TGAGACTCTG TCTCAAAAAT AAATAAATAA ATAAATAAAT AAATAAATAA ATAAATAAAT 60
AAATAAATAA GTAAAAAGA AAGATGC 88

(2) INFORMATION FOR SEQ ID NO:263:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

-319-

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

GTGTGTGTGT GTGTGTGTT TTTTTTTAAC AGGGATTGG GGAATTATTT GAGA

54

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

TTCCCCAAAT CCCTGTATAA AAC

23

(2) INFORMATION FOR SEQ ID NO:265:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

TTCCCCAAAT CCCTGTATAA AAAAC

25

(2) INFORMATION FOR SEQ ID NO:266:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

-320-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

TTCCCCAAAT CCCTGTTAAA AAAAAAC

27

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

GTAAAACGAC CGCCAGTGCC AAGCTTGCAT GCCTGCAGGT CGACTCTAGA GGATCCCCGG 60
GTACCGAGCT CGAATTCGTA ATCATGGTCA TAGCTGTTTC CTG 103

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GAGTCAGGTG CGCCATGCCT CAAACAGACA CCATGGTGCA CCTGACTCCT GAGGAGNCTG 60
GGCATGTGGA GACAGAGA 78

(2) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GACTCAGGTG CACATGGTGT CTGTTTGAGG 60

-321-

CATGGCGCAC CTGAGCTC

78

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CGCCATGGTG TCTGTTTGAG 60
GCATGGCGCA CGTGACTC 78

(2) INFORMATION FOR SEQ ID NO:271:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CGCCATGGTG TCTGTTTGAG 60
GCATGGCGCA CCTGACTCCT GA 82

(2) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CG 42

(2) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

-322-

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

CACCTGACTC CTA

13

(2) INFORMATION FOR SEQ ID NO:274:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

CACCTGACTC CTGGA

14

(2) INFORMATION FOR SEQ ID NO:275:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

CACCTGACTC CTGA

14

(2) INFORMATION FOR SEQ ID NO:276:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-323-

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

CCATGGTGTC TGTTTGAGGC ATGGCG

26

(2) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCAGAG CATTAAAGAT 60
AGTCATCTTG GGGCT 75

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

ACCTAGCGTT CAGTTCGACT GAGATAATAC GACTCACTAT AGCAGCTCTC ATTTTCATA 60
C 61

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

-324-

GUCACUACAG GUGAGCUCCA

20

(2) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

CTCAGTCCAC GTGGTACCCT GCTG

24

(2) INFORMATION FOR SEQ ID NO:281:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

CATTGCTTC TGACACAACT GTGTTCACTA GCAACCTCAA ACAGACACCA TGGTGACCT
GACTCCTGAG GAGAAGTCTG CCGTT 85

(2) INFORMATION FOR SEQ ID NO:282:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

ACGGGTCCCG GAGTGGTGTC GC

22

(2) INFORMATION FOR SEQ ID NO:283:

-325-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGGTTG 60
GTATCAAGGT TACAAG 76

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGATTG 60
GTATCAAGGT TACAAG 76

(2) INFORMATION FOR SEQ ID NO:285:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGGTTG 60
CTATCAAGGT TACAAG 76

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-326-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGCAGGTTG 60
GCATCAAGGT TACAAG 76

(2) INFORMATION FOR SEQ ID NO:287:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

ACAGGTTTAA GGAGACCAAT AGAACTGGG CATGTGGAGA CAGAGAAG 48

(2) INFORMATION FOR SEQ ID NO:288:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

GACGACGACT GCTACCTGAC TCCA 24

(2) INFORMATION FOR SEQ ID NO:289:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-327-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

ACAGCGCACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:290:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

TGGAGTCAGG TAGCAGTC

18

(2) INFORMATION FOR SEQ ID NO:291:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTCCAGA CATTAAAGAT 60

(2) INFORMATION FOR SEQ ID NO:292:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 65 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

AGTCATCTTG GGGCTGTCGA GAGTAAAGG TATGTCAGTC ATAGTTAAGA CCTTCTTAAAGGTCT

65

-328-

(2) INFORMATION FOR SEQ ID NO:293:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

GTAATTTCTA TCAGTAGAAC CCCGA

25

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCCAGA CATTAAAGAT 60

(2) INFORMATION FOR SEQ ID NO:295:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

AGTCATCTTG GGGCT

15

(2) INFORMATION FOR SEQ ID NO:296:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-329-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCAGAG CATTAAAGAT 60

(2) INFORMATION FOR SEQ ID NO:297:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

AGTCATCTTG GGGCTA

16

(2) INFORMATION FOR SEQ ID NO:298:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

CAGCTCTCAT TTTCCATACA TTAAAGATAG TCATCTTGGG GCT

43

(2) INFORMATION FOR SEQ ID NO:299:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

-330-

CAGCTCTCAT TTTCCATACA TTAAAGATAG TCATCTTGGG GCTA

44

(2) INFORMATION FOR SEQ ID NO:300:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

CAGCTCTCAT TTTCCATACA GT

22

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

CAGCTCTCAT TTTCCATACA T

21

(2) INFORMATION FOR SEQ ID NO:302:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

GCCTGGTACA CTGCCAGGCG CTTCTGCAGG TCATCGGCAT CGCGGAGGAG

50

(2) INFORMATION FOR SEQ ID NO:303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid

-331-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

GCCTGGTACA CTGCCAGGCA CTTCTGCAGG TCATCGGCAT CGCGGAGGAG

50

- (2) INFORMATION FOR SEQ ID NO:304:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

GATGCCGATG ACCTGCAGAA G

21

- (2) INFORMATION FOR SEQ ID NO:305:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

GATGCCGATG ACCTGCAGAA GC

22

- (2) INFORMATION FOR SEQ ID NO:306:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-332-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

GATGCCGATG ACCTGCAGAA GTGC

24

(2) INFORMATION FOR SEQ ID NO:307:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

GATGATCCGA CG

12

(2) INFORMATION FOR SEQ ID NO:308:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

CTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:309:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

GATGATCCGA CG

12

-333-

(2) INFORMATION FOR SEQ ID NO:310:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

GGCGCGGACA TGGAGGACGT GTGCGGCCGC CTGGT

35

(2) INFORMATION FOR SEQ ID NO:311:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

TCCGCGATGC CGATGACCTG CAGAAGCGCC TGGC

34

(2) INFORMATION FOR SEQ ID NO:312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

CGGCTGCGAT CACCGTGCGG CACAGCT

27

(2) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-334-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

CGGCTGCGAT CACCGTGCGG T

21

(2) INFORMATION FOR SEQ ID NO:314:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

CGGCTGCGAT CACCGTGCGG AACAGCT

27

(2) INFORMATION FOR SEQ ID NO:315:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

CGGCTGCGAT CACCGTGCGG CA

22

(2) INFORMATION FOR SEQ ID NO:316:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-335-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

CGGCTGCGAT CACCGTGCGG TA

22

(2) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

CGGCTGCGAT CACCGTGCGG A

21

(2) INFORMATION FOR SEQ ID NO:318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

ATCATCAACT GGAAGATCAG GTCAGGAGCC ACTTGCCANC CT

42

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

ATCATCACAC TGGAAGACTC CAGGTCAGGA GCC

33

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

-336-

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

ATCCACTACA ACTACATGTG TAACAGTTGG wGCwwGCC

48

-337-

WHAT IS CLAIMED IS:

1. A process for determining the sequence of a target nucleic acid molecule comprising the steps of:

- 5 a) generating at least two nucleic acid fragments from the target nucleic acid; and
- b) analyzing the at least two fragments by a mass spectrometry format, and thereby determine the sequence of the target nucleic acid molecule.

10 2. A process of claim 1, wherein in step a), an endonuclease is contacted with the target nucleic acid to generate the at least two nucleic acid fragments.

 3. A process of claim 2, wherein the endonuclease is a restriction enzyme that can recognize and cleave at least one restriction site in the target nucleic acid.

15 4. A process of claim 2, wherein the target nucleic acid is a deoxyribonucleic acid and the nuclease is a deoxyribonuclease.

 5. A process of claim 2, wherein the target nucleic acid is a ribonucleic acid and the nuclease is a ribonuclease.

20 6. A process of claim 5, wherein the ribonuclease is selected from the group consisting of: the G-specific T₁ ribonuclease, the A-specific U₂ ribonuclease, the A/U specific PhyM ribonuclease, the U/C specific ribonuclease A, the C-specific chicken liver ribonuclease and crisavitin.

25 7. A process of claim 1, wherein in step a), nucleic acid fragments are generated by performance of a combined amplification and base-specific termination reaction.

 8. A process of claim 7, wherein the combined amplification and base-specific termination reaction is performed using a first polymerase, which has a relatively low affinity towards at least one chain terminating

-338-

nucleotide and an a second polymerase, which has a relatively high affinity towards at least one chain terminating nucleotide.

9. A process of claim 8, wherein the first and second polymerases are thermostable DNA polymerases.

5 10. A process of claim 9, wherein the thermostable DNA

polymerases are selected from the group consisting of: Taq DNA polymerase, AmpliTaq FS DNA polymerase, Deep Vent (exo-) DNA polymerase, Vent DNA polymerase, Vent (exo-) DNA polymerase, Vent DNA polymerase, Vent (exo-) DNA polymerase, Deep Vent DNA

10 polymerase, Thermo Sequenase, exo(-) *Pseudococcus furiosus* (Pfu) DNA polymerase, AmpliTaq, Ultman, 9 degree Nm, Tth, Hot Tub, *Pyrococcus furiosus* (Pfu) and *Pyrococcus woesei* (Pwo) DNA polymerase.

11. A process of claim 1, wherein the at least two nucleic acid fragments generated in step a) include mass modified nucleotides.

15 12. A process of claim 1, wherein the at least two fragments comprise a 3' tag.

13. A process of claim 1, wherein the at least two fragments comprise a 5' tag.

20 14. A process of claim 12 or 13, wherein the tag is a non-natural tag.

15. A process of claim 14, wherein the non-natural tag is selected from the group consisting of: an affinity tag and a mass marker.

16. A process of claim 15, wherein the affinity tag facilitates immobilization of the nucleic acid to a solid support.

25 17. A process of claim 16, wherein the affinity tag is biotin or a nucleic acid sequence that is capable of binding to a capture nucleic acid sequence that is bound to a solid support.

-339-

18. A process of claim 1, wherein the process additionally comprises the step of: ordering the at least two nucleic acid fragments to determine the sequence of the target nucleic acid.

5 19. A process for detecting a target nucleic acid present in a biological sample, comprising the steps of:

- 10 a) performing on a nucleic acid obtained from a biological sample; a first polymerase chain reaction using a first set of primers, which are capable of amplifying a portion of the nucleic acid containing the target nucleic acid, thereby producing a first amplification product; and
- b) detecting the first amplification product by mass spectrometry, wherein detection of the target nucleic acid indicates that the target nucleic acid is present in the biological sample.

15 20. A process of claim 19, wherein prior to step b), a second polymerase chain reaction is performed on the first amplification product using a second set of primers, which are capable of amplifying at least a portion of the first amplification product, which contains the target nucleic acid.

20 21. A process of claim 19 or 20, wherein prior to step b), the target nucleic acid is immobilized to a solid support.

22. A process of claim 21, wherein the target nucleic acid is reversibly immobilized.

25 23. A process of claim 22, wherein the target nucleic acid can be cleaved from the solid support by a chemical, enzymatic or physical process.

24. A process of claim 23, wherein immobilization is accomplished via a photocleavable bond.

25. A process of claim 22, wherein the target nucleic acid is cleaved from the support during step b).

-340-

26. A process of claim 21, wherein the solid support is selected from the group consisting of: beads, flat surfaces, chips, capillaries, pins, combs and wafers.

5 27. A process of claim 21, wherein immobilization is accomplished by hybridization between a complementary capture nucleic acid molecule immobilized to a solid support, and a portion of the nucleic acid molecule, which is distinct from the target nucleic acid sequence.

28. A process of claim 19 or 20, wherein prior to step b), the target nucleic acid is purified.

10 29. A process of claim 19 or 20, wherein the primer or first or second amplification product is conditioned.

30. A process of claim 29, wherein the primer or first or second amplification product is conditioned by phosphodiester backbone modification.

15 31. A process of claim 30, wherein the phosphodiester backbone modification is a cation exchange.

32. A process of claim 29, wherein the primer or first or second amplification product is conditioned by contact with an alkylating agent or trialkylsilyl chloride.

20 33. A process of claim 29, wherein conditioning is effected by including at least one nucleotide that reduces sensitivity for depurination in the primer or first or second amplification product.

34. A process of claim 33, wherein the nucleotide is an N7- or N9- deazapurine nucleotide or 2' fluoro 2' deoxy nucleotide.

25 35. A method for detecting neoplasia/malignancies in a tissue or cell sample, comprising detecting telomerase activity, mutation of a proto-oncogene, expression of a tumor specific gene in the sample by detecting nucleic acids that encode the telomerase, that are specific for the mutation or that encode the tumor-specific by mass spectrometry.

-341-

36. The method of claim 35 that is a method for detecting neoplasia/malagnancies in a tissue or cell sample, comprising:

- 5
- a) isolating telomerase from the sample and adding a synthetic DNA primer, which is optionally immobilized, complementary to a telomeric repeat, and all four deoxynucleotide triphosphates under conditions that result in telomerase specific extension of the synthetic DNA;
- 10
- b) amplifying the telomerase extended DNA product; and
- c) detecting the DNA product by mass spectrometry, wherein telomerase-specific extension is indicative of neoplasia/malignancy.

37. The method of claim 36, wherein the primer contains a
15 linker moiety for immobilization on a support; and the amplified primers are isolated conjugating the linker portion to a solid support.

38. The method of claim 35 that is a method for identifying transformed cells or tissues, comprising:

- 20
- a) in a cell or tissue sample, amplifying a portion of a proto-oncogene that includes a codon indicative of transformation, wherein one primer comprises a linker moiety for immobilization;
- c) immobilizing DNA via the linker moiety to a solid support, optionally in the form of an array;
- 25
- d) hybridizing a primer complementary to the proto oncogene sequence that is upstream from the codon
- e) adding 3dNTPs/1 ddNTP and DNA polymerase and extending the hybridized primer to the next ddNTP location;

-342-

- f) ionizing/volatizing the sample; and
- g) detecting the mass of the extended DNA, whereby mass indicates the presence of wild-type or mutant alleles. The presence of a mutant allele at the codon is diagnostic for neoplasia.

5

39. The method of claim 38, wherein the proto-oncogene is the RET-proto-oncogene.

40. The method of claim 35 that is a method for detecting expression of a tumor-specific gene, comprising:

10

- a) isolating polyA RNA from the sample;
- c) preparing a cDNA library using reverse transcription;
- d) amplifying a cDNA product, or portion thereof, of the tumor-specific gene, wherein one oligo primer comprises a linker moiety;

15

- e) isolating the amplified product by immobilizing the DNA to a solid support via the linker moiety;
- f) optionally conditioning the DNA;
- g) ionizing/volatizing sample and detecting the presence

of a DNA peak that is indicative of expression of the gene.

20

41. The method of claim 40, wherein the cells are bone marrow cells, the gene is the tyrosine hydroxylase gene, and expression of the gene is indicative of neuroblastoma.

42. A method for directly detecting a double-stranded nucleic acid using matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF) mass spectrometry, comprising:

25

- a) isolating a double-stranded DNA fragment from a cell or tissue sample;

-343-

- 5 b) preparing the double-stranded DNA for analysis under conditions that increase the ratio of dsDNA:ssDNA, wherein the conditions include one or all of the following: preparing samples for analysis at reduced temperatures (i.e. 4 ° C), and using of higher DNA concentrations in the matrix to drive duplex formation;
- c) ionizing/volatizing the sample of step b), wherein low acceleration voltage of the ions are used;
- d) detecting the presence of the double-stranded DNA.
- 10 43. A method for comparing DNA samples to discern relatedness or to detect mutations, comprising:
- a) obtaining biological a plurality of samples;
- b) amplifying a region of DNA from each sample that contains two or more microsatellite DNA repeat sequences;
- 15 c) ionizing/volatizing the amplified DNA;
- d) detecting the presence of the amplified DNA and comparing the molecular weight of the amplified DNA, wherein different sizes are indicative of non-identity between or among the samples.
- 20 44. The method of claim 43, wherein non-identity is indicative of the presence of a mutation in the DNA in one sample, non-relatedness or non-HLA compatibility between or among the individuals from whom the samples were obtained.
45. The method of claim 43 or 44, wherein a plurality of
- 25 markers are examined simultaneously.

-344-

46. A method for detecting a target nucleic acid in a sample, comprising:

- a) amplifying a target nucleic acid sequence using;
 - (i) a first primer, wherein:
 - 5 the 5'-end shares identity to a portion of the target DNA immediately downstream from the targeted codon followed by a sequence that introduces a unique restriction endonuclease site, and
 - the 3'-end primer is self-complementary; and
 - 10 (ii) a second downstream primer that contains a tag;
- b) immobilizing the double-stranded amplified DNA to a solid support via the linker moiety;
- c) denaturing the immobilized DNA and isolating the non-immobilized DNA strand;
- 15 d) annealing the intracomplementary sequences in the 3'-end of the isolated non-immobilized DNA strand, such that the 3'-end is extendable by a polymerase;
- f) extending the annealed DNA by adding DNA polymerase, 3 dNTPs/1 ddNTP;
- 20 g) cleaving the extended double stranded stem loop DNA with the unique restriction endonuclease and removing the cleaved stem loop DNA;
- i) ionizing/volatizing the extended product; and
- j) detecting the presence of the extended target nucleic acid, where
25 by the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation at the target codon(s).

-345-

47. A method for detecting a target nucleic acid in a biological sample using RNA amplification, comprising:

amplifying the target nucleic acid using a primer comprising a region complementary to the target sequence and a region that encodes
5 a promoter;

synthesizing RNA using an RNA polymerase that recognizes the promoter;

detecting the resulting RNA using mass spectrometry.

48. A primer for mass spectrometric analyses, comprising all or
10 at least about 20, preferably about 16, bases of any of the sequence of nucleotide sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95, 98, 101-110, 112-123, 126, 128 and 129, wherein the primer is unlabeled.

49. The primer of claim 48, further comprising a mass
15 modifying moiety.

50. A process for detecting a target nucleic acid sequence present in a biological sample, comprising the steps of:

a) obtaining a nucleic acid molecule containing a target nucleic acid sequence from a biological sample;

20 b) immobilizing the target sequence on the support via thiol linkages, whereby the target is present at a sufficient density to detect it using mass spectrometry;

c) hybridizing a detector oligonucleotide with the target nucleic acid sequence;

25 d) removing unhybridized detector oligonucleotide;

e) ionizing and volatilizing the product of step c); and

f) detecting the detector oligonucleotide by mass spectrometry, wherein detection of the detector oligonucleotide indicates the presence of the target nucleic acid sequence in the biological sample.

-346-

51. The process of claim 50, wherein the target nucleic acid molecule is amplified prior to immobilization.

52. The process of claim 50 or 52, wherein at least one of the detector oligonucleotide or the target nucleic acid sequence has been
5 conditioned.

53. A process of any of claims 50-52, wherein the solid support is selected from the group consisting of: beads, flat surfaces, pins and combs.

54. A process of any of claims 50-53, wherein target nucleic
10 acid is immobilized in the form of an array.

55. A process of any of claims 50-54, wherein the support is a silicon wafer.

56. A process of any of claims 51-55, wherein the target nucleic acid molecule is amplified by an amplification procedure selected
15 from the group consisting of cloning, transcription, the polymerase chain reaction (PCR), the ligase chain reaction (LCR), and strand displacement amplification (SDA).

57. A process of any of claims 50-56, wherein the mass spectrometer is selected from the group consisting of: Matrix-Assisted
20 Laser Desorption/Ionization Time-of-Flight (MALDI-TOF), Electrospray (ES), Ion Cyclotron Resonance (ICR), and Fourier Transform.

58. A process of any of claims 50-57, wherein the sample is conditioned by mass differentiating at least two detector oligonucleotides or oligonucleotide mimetics to detect and distinguish at least two target
25 nucleic acid sequences simultaneously.

59. A process of claim 58, wherein the mass differentiation is achieved by differences in the length or sequence of the at least two oligonucleotides.

-347-

60. A process of claim 59, wherein the mass differentiation is achieved by the introduction of mass modifying functionalities in the base, sugar or phosphate moiety of the detector oligonucleotides.

61. A process of claim 58, wherein the mass differentiation is
5 achieved by exchange of cations at the phosphodiester bond.

62. A process of any of claims 50-61, wherein the nucleic acid molecule obtained from a biological sample is amplified into DNA using mass modified dideoxynucleoside triphosphates and DNA dependent DNA polymerase prior to mass spectrometric detection.

10 63. A process of any of claims 50-62, wherein the nucleic acid molecule obtained from a biological sample is amplified into RNA using mass modified ribonucleoside triphosphates and DNA dependent RNA polymerase prior to mass spectrometric detection.

64. A process of any of claims 50-63, herein the target nucleic
15 acid sequence is indicative of a disease or condition selected from the group consisting of a genetic disease, a chromosomal abnormality, a genetic predisposition, a viral infection, a fungal infection and a bacterial infection.

65. A method of determining a sequence of a nucleic acid,
20 comprising the steps of:

- (i) obtaining multiple copies of the nucleic acid to be sequenced;
- (ii) cleaving the multiple copies from a first end to a second end with an exonuclease to sequentially release individual nucleotides;
- (iii) identifying each of the sequentially released nucleotides by

25 mass spectrometry; and

- (iv) determining the sequence of the nucleic acid from the identified nucleotides, wherein the nucleic acid is immobilized by covalent attachment to a solid support via at least one sulfur atom.

-348-

66. A method of determining a sequence of a nucleic acid, comprising the steps of:

- (i) obtaining multiple copies of the nucleic acid to be sequenced;
- (ii) cleaving the multiple copies from a first end to a second end
- 5 with an exonuclease to produce multiple sets of nested nucleic acid fragments;
- (iii) determining the molecular weight value of each one of the sets of nucleic acid fragments by mass spectrometry; and
- (iv) determining the sequence of the nucleic acid from the
- 10 molecular weight values of the sets of nucleic acid fragments, wherein the nucleic acid is immobilized by covalent attachment to a solid support via at least one sulfur atom.

67. The process of claim 65 or 66, wherein the nucleic acids are covalently bound to a surface of the support at a density of at least 20

15 fmol/mm².

68. The method of any of claims 50-67, wherein immobilization is effected by a method comprising:

- reacting a thiol-containing insoluble support with a nucleic acid comprising a thiol-reactive group under conditions such that a covalent
- 20 bond is formed;
- thereby immobilizing the nucleic acid on the insoluble support.

69. The method of claim 68, further including the step of modifying the insoluble support with a thiol-containing reagent, to form a thiol-containing insoluble support.

25 70. The method of claim 68 or 69, wherein the thiol-reactive cross-linking reagent is N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB).

71. The method of claim 65 or claim 66, wherein the nucleic acid is a 2'-deoxyribonucleic acid (DNA).

-349-

72. The method of claim 65 or claim 66, wherein the nucleic acid is a ribonucleic acid (RNA).

73. The method of any of claims 65-71, wherein the exonuclease is selected from the group consisting of snake venom
5 phosphodiesterase, spleen phosphodiesterase, Bal-31 nuclease, *E. coli* exonuclease I, *E. coli* exonuclease VII, Mung Bean Nuclease, S1 Nuclease, an exonuclease activity of *E. coli* DNA polymerase 1, an exonuclease activity of a Klenow fragment of DNA polymerase 1, an exonuclease activity of T4 DNA polymerase, an exonuclease activity of
10 T7 DNA polymerase, an exonuclease activity of Taq DNA polymerase, an exonuclease activity of DEEP VENT DNA polymerase, *E. coli* exonuclease III, lambda exonuclease and an exonuclease activity of VENT_RDNA polymerase.

74. The method of any of claims 65-74, wherein the nucleic acid
15 comprises mass-modified nucleotides.

75. The method of claim 74, wherein the mass-modified nucleotides modulate the rate of the exonuclease activity.

76. The method of claim 74, wherein the sequentially released nucleotides are mass-modified subsequent to exonuclease release and
20 prior to mass spectrometric identification.

77. The method of claim 76, wherein the sequentially released nucleotides are mass-modified by contact with an alkaline phosphatase.

78. A method of any of claims 65-77, wherein the mass spectrometry format is matrix assisted laser desorption (MALDI) mass
25 spectrometry or electrospray (ES) mass spectrometry.

79. A method of any of claims 65-79, wherein immobilization is effected by a method, comprising:

-350-

reacting the surface of the substrate with a solution of 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface of the substrate; and

derivatizing the surface of a substrate with iodoacetamido
5 functionalities by reacting the uniform layer of primary amines with a solution of N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB).

80. A primer, comprising all least about 20, preferably about 16, bases of any of the sequence of nucleotides sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95, 98, 101-110, 112-123,
10 126, 128 and 129.

81. The primers of claim 80 that is unlabeled, and optionally includes a mass modifying moiety, which is preferably attached to the 5'end.

82. The method of any of claims 1-79, wherein nucleic acid is
15 immobilized to a solid support via a selectively cleavable linker.

83. The method of claim 82, wherein the linker is thermocleavable, enzymatically cleavable, photocleavable or chemically cleavable.

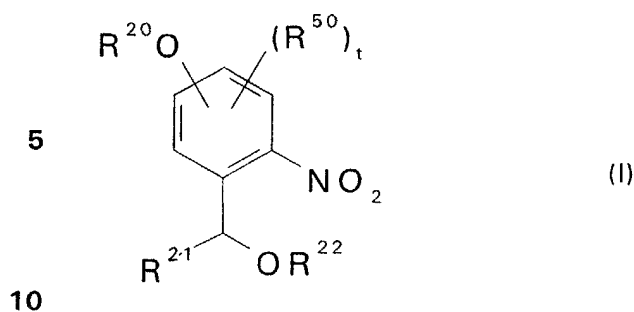
82. The method of claim 82, wherein the linker is a trityl linker.

20 83. The method of claim 82, wherein the linker is selected from the group consisting of 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)-phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane and 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

25 84. A photolabile linker, comprising a compound of formula:

30

-351-



wherein:

R^{20} is selected from the group consisting of ω -(4,4'-dimethoxytrityloxy)alkyl and ω -hydroxyalkyl;

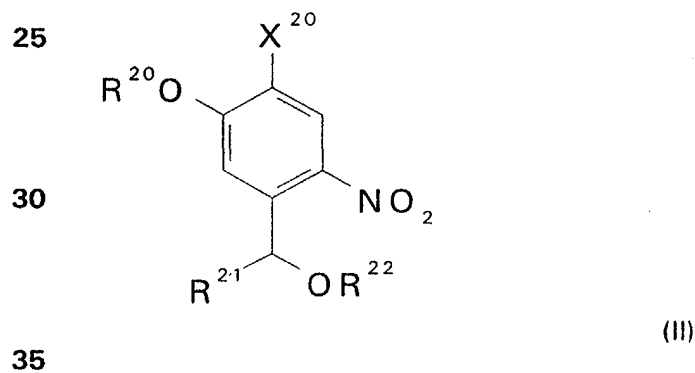
15 R^{21} is selected from the group consisting of hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxycarbonyl and carboxy;

R^{22} is selected from the group consisting of hydrogen and (dialkylamino)(ω -cyanoalkoxy)P-;

t is 0-3; and

20 R^{50} is selected from the group consisting of alkyl, alkoxy, aryl and aryloxy.

85. The photocleavable linker of claim 84, wherein the linkers are of formula II:



-352-

wherein:

R^{20} is selected from the group consisting of ω -(4,4'-dimethoxytrityloxy)alkyl, ω -hydroxyalkyl and alkyl;

5 R^{21} is selected from the group consisting of hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxy carbonyl and carboxy;

R^{22} is selected from the group consisting of hydrogen and (dialkylamino)(ω -cyanoalkoxy)P-; and

10 X^{20} is selected from the group consisting of hydrogen, alkyl or OR²⁰.

86. The photocleavable linker of claim 85, wherein:

R^{20} is selected from the group consisting of 3-(4,4'-dimethoxytrityloxy)propyl, 3-hydroxypropyl and methyl;

15 R^{21} is selected from the group consisting of hydrogen, methyl and carboxy;

R^{22} is selected from the group consisting of hydrogen and (diisopropylamino)(2-cyanoethoxy)P-; and

X^{20} is selected from the group consisting of hydrogen, methyl or OR²⁰.

20 87. The photocleavable linker of claim 85, wherein:

R^{20} is 3-(4,4'-dimethoxytrityloxy)propyl;

R^{21} is methyl;

R^{22} is (diisopropylamino)(2-cyanoethoxy)P-; and

X^{20} is hydrogen.

25 88. The photocleavable linker of claim 86, wherein:

R^{20} is methyl;

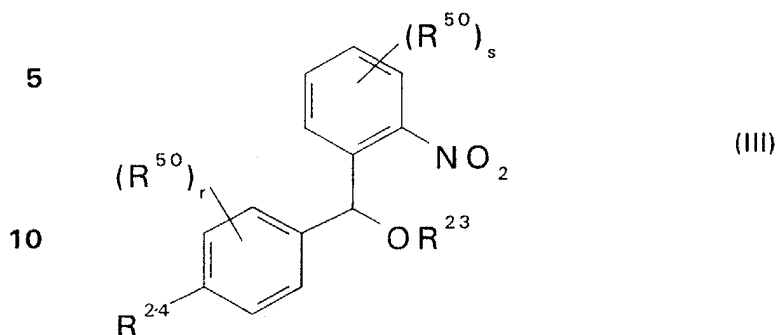
R^{21} is methyl;

R^{22} is (diisopropylamino)(2-cyanoethoxy)P-; and

X^{20} is 3-(4,4'-dimethoxytrityloxy)propoxy.

-353-

88. A photocleavable linker, comprising a compound of formula III:



15 wherein:

R^{23} is selected from the group consisting of hydrogen and (dialkylamino)(ω -cyanoalkoxy)P-;

R^{24} is selected from ω -hydroxyalkoxy, ω -(4,4'-dimethoxytrityloxy)alkoxy, ω -hydroxyalkyl and ω -(4,4'-dimethoxytrityloxy)alkyl, and is unsubstituted or substituted on the alkyl or alkoxy chain with one or more alkyl groups;

r and s are each independently 0-4; and

R^{50} is alkyl, alkoxy, aryl or aryloxy.

89. The photocleavable linker of claim 88, wherein:

25 R^{24} is ω -hydroxyalkyl or ω -(4,4'-dimethoxytrityloxy)alkyl, and is substituted on the alkyl chain with a methyl group.

90. The photocleavable linker of claim 88, wherein:

R^{23} is selected from the group consisting of hydrogen and (diisopropylamino)(2-cyanoethoxy)P-; and

30 R^{24} is selected from the group consisting of 3-hydroxypropoxy, 3-(4,4'-dimethoxytrityloxy)propoxy, 4-hydroxybutyl, 3-hydroxy-1-propyl, 1-hydroxy-2-propyl, 3-hydroxy-2-methyl-1-propyl, 2-hydroxyethyl, hydroxymethyl, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-

-354-

dimethoxytrityloxy)-1-propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-dimethoxytrityloxymethyl.

91. The photocleavable linker of claim 90, wherein r and s are
5 both 0.

92. The photocleavable linker of claim 91, wherein:

R²³ is (diisopropylamino)(2-cyanoethoxy)P-; and

R²⁴ is selected from the group consisting of 3-(4,4'-dimethoxytrityloxy)propoxy, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-
10 dimethoxytrityloxymethyl.

93. The photocleavable linker of claim 92, wherein:

R²⁴ is 3-(4,4'-dimethoxytrityloxy)propoxy.

15 94. The photocleavable linker of claim 84, where in the linker is selected from the group consisting of 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane and 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

1 / 123

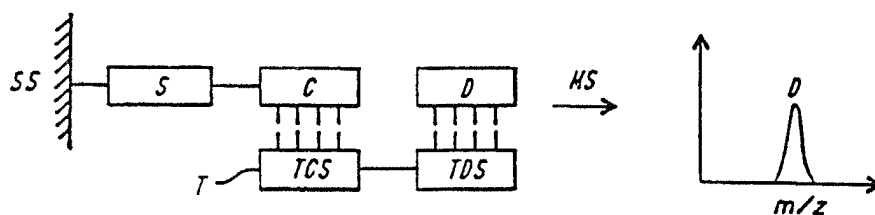


FIG. 1A

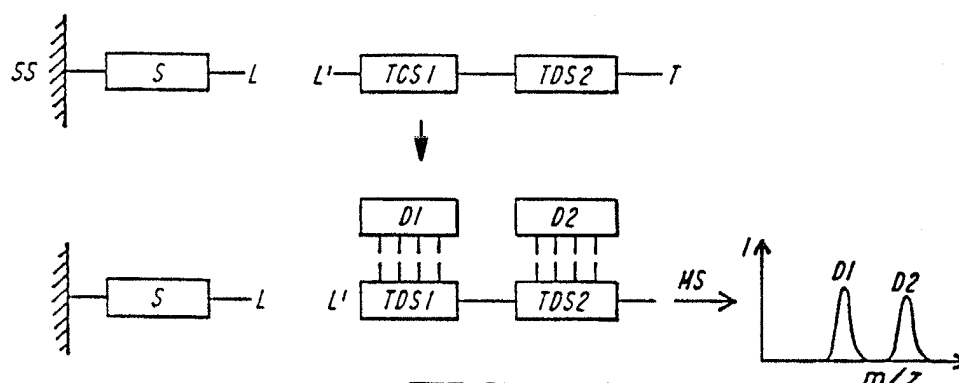


FIG. 1B

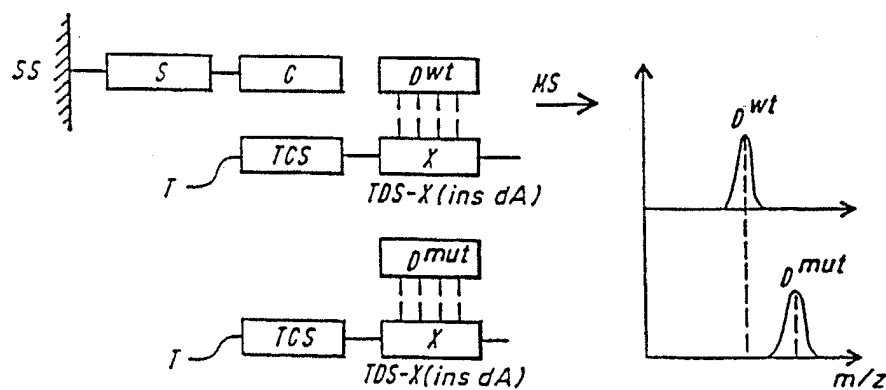
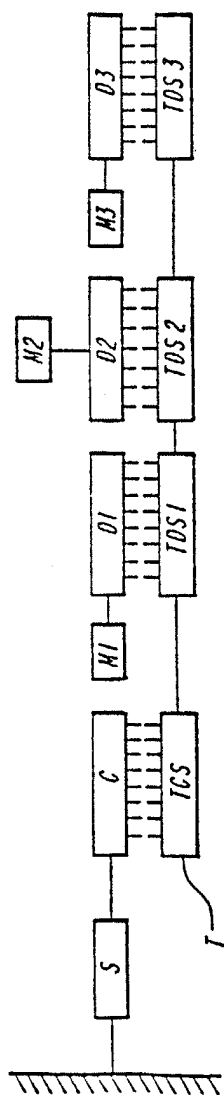
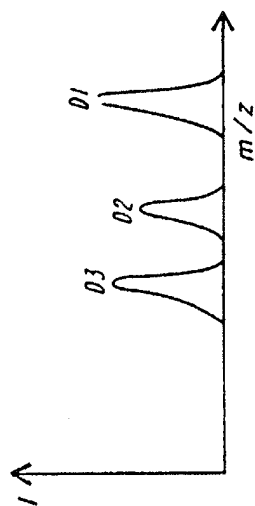


FIG. 1C

2/123



MS



$I = \text{INTENSITY}$
 $m/z = \frac{\text{MASS}}{\text{CHARGE}}$

$H1 > H2 > H3$

D1 - D3 SIMILAR MOLECULAR WEIGHT

H : MASS-MODIFYING FUNCTION

— : POLYMERIC SUPPORT

FIG. 2

3/123

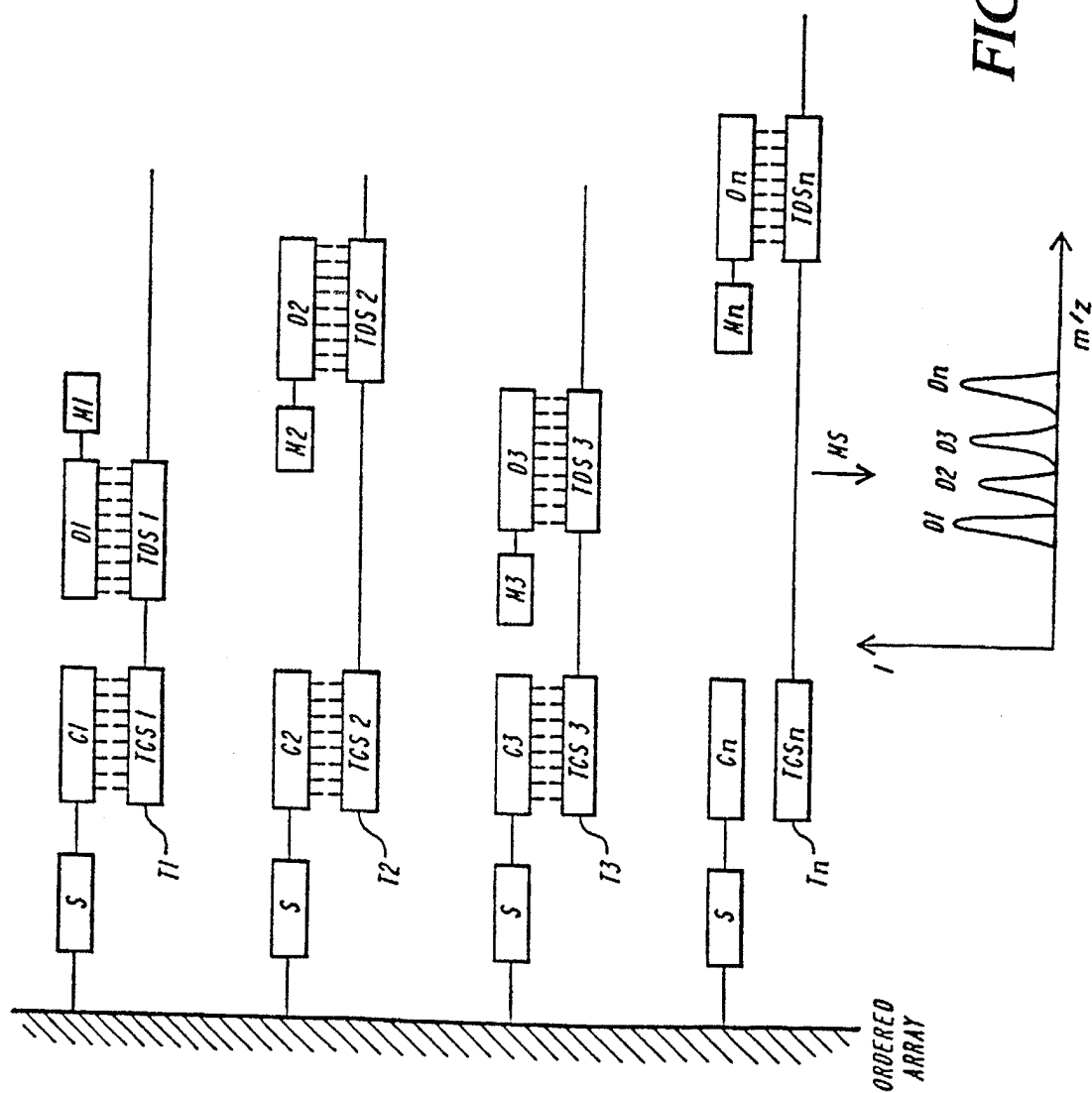


FIG. 3

4/123

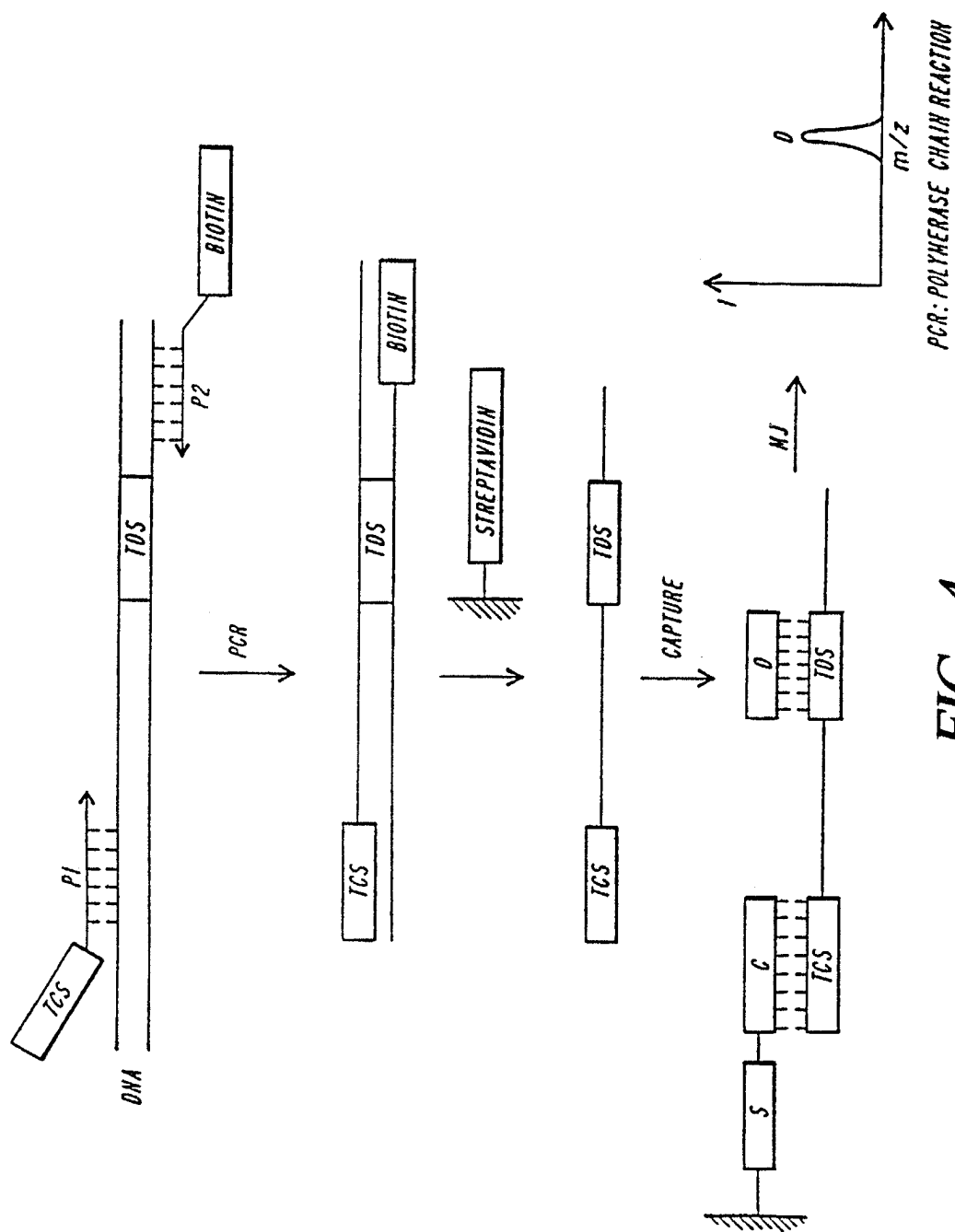
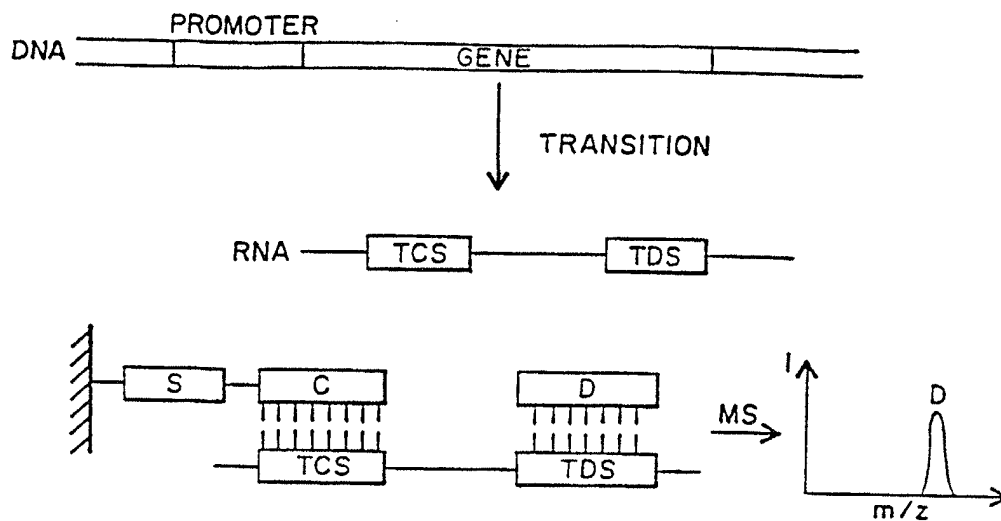
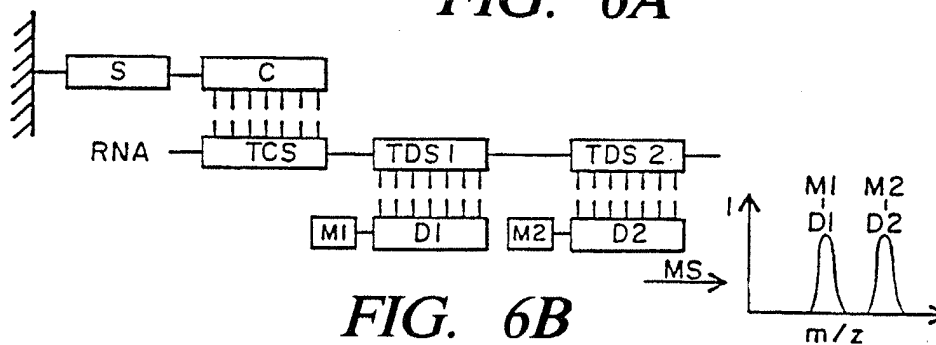
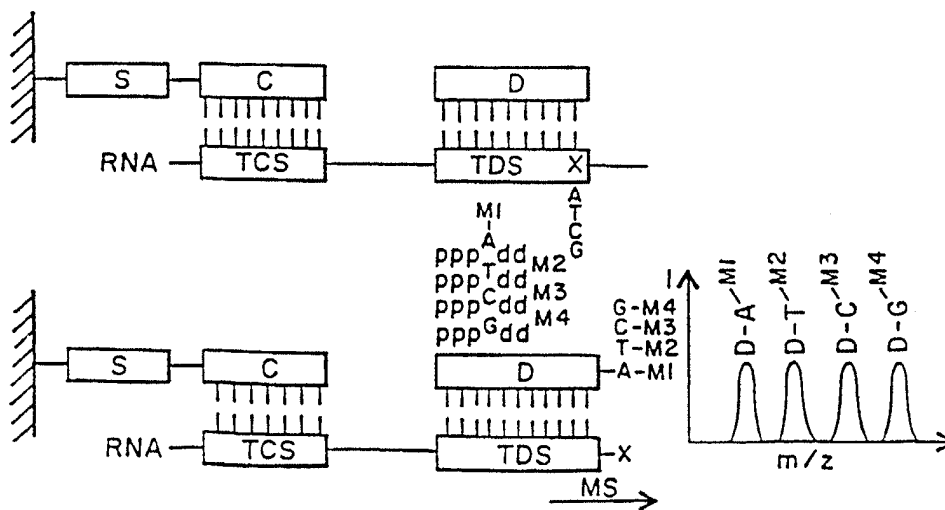
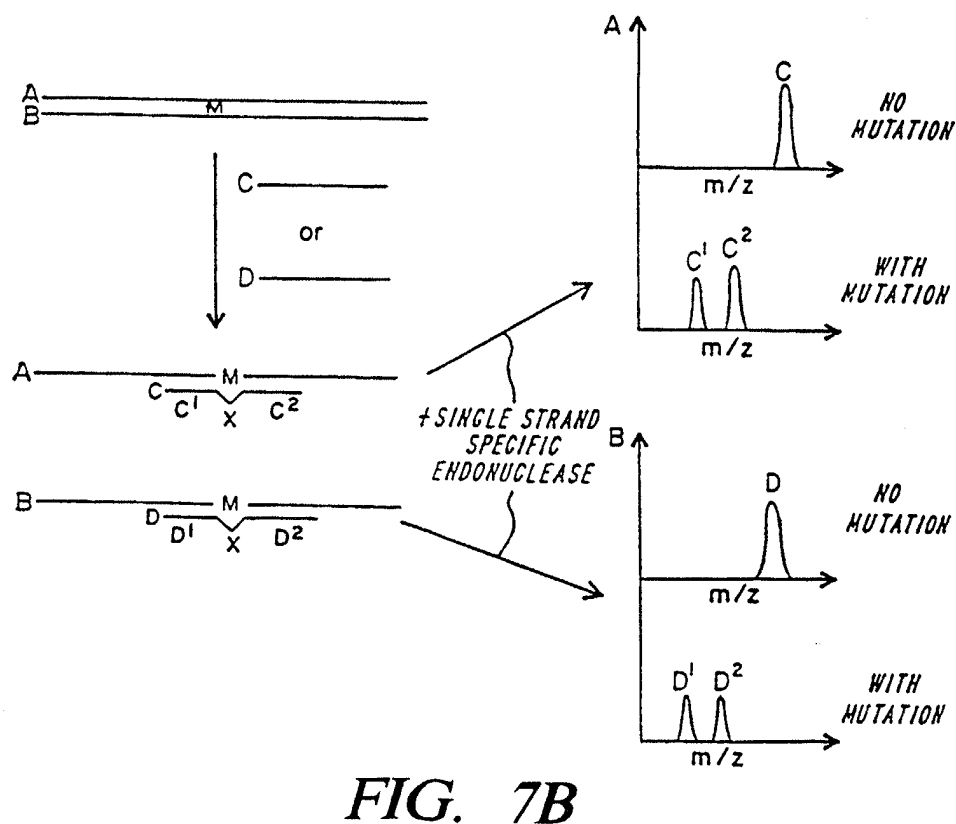
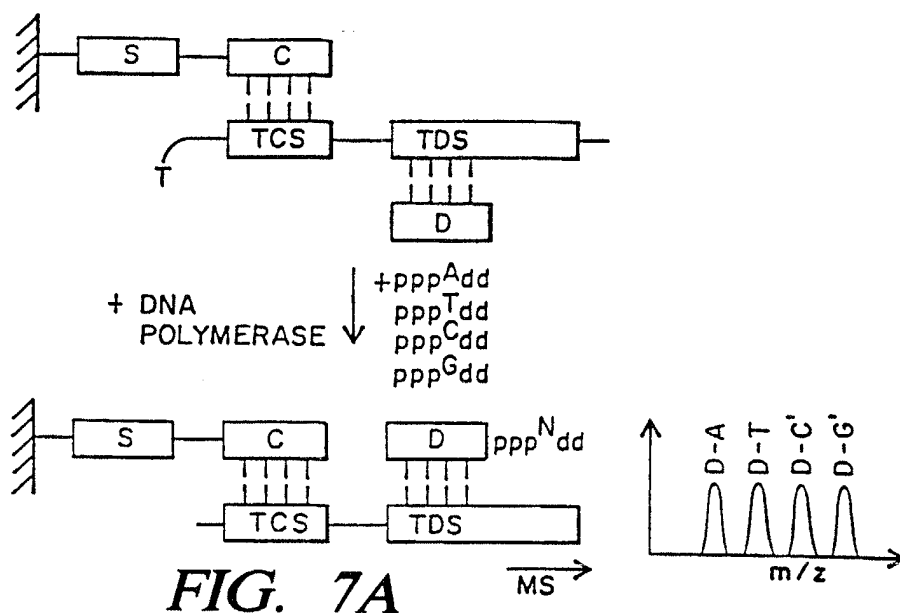


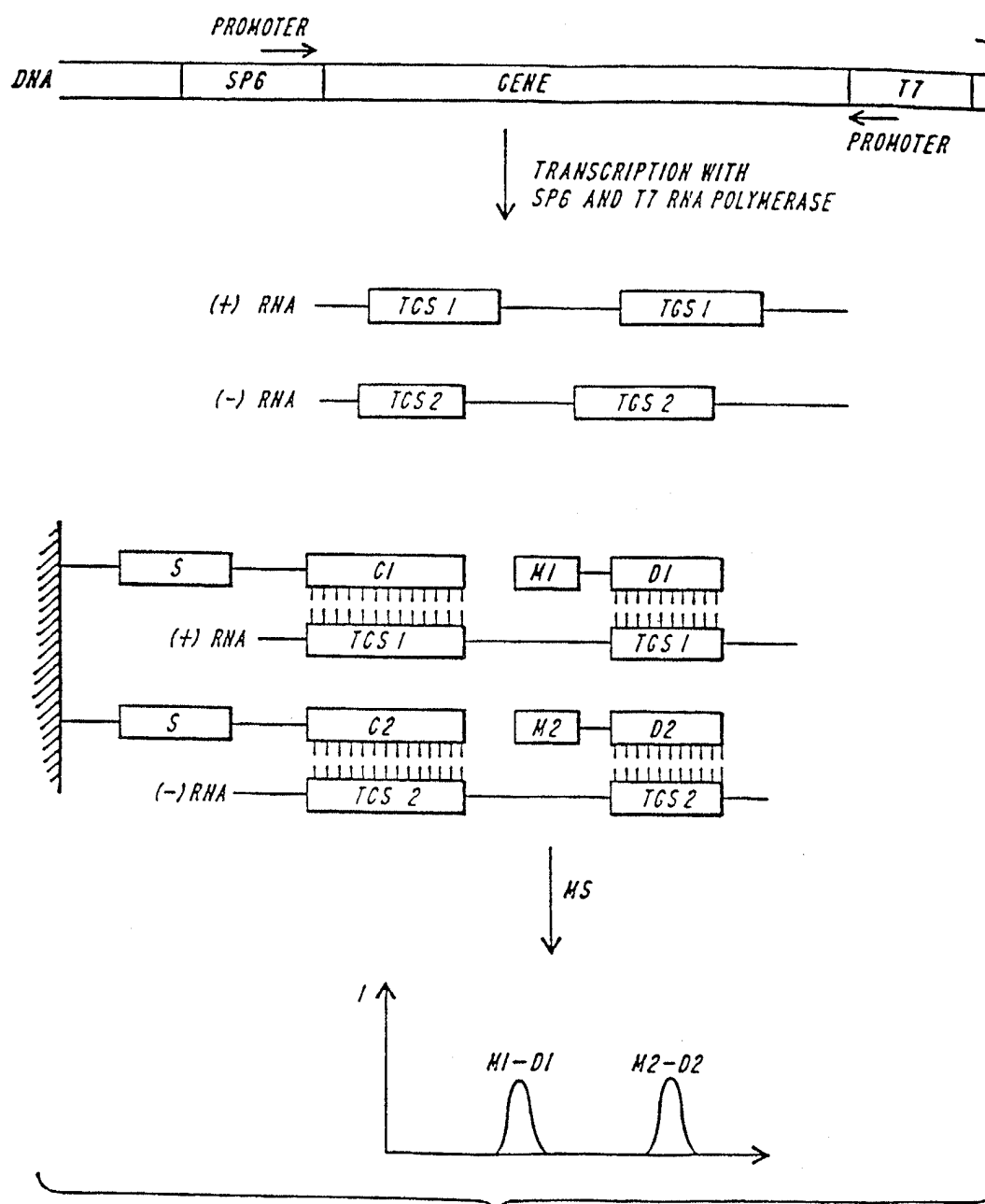
FIG. 4

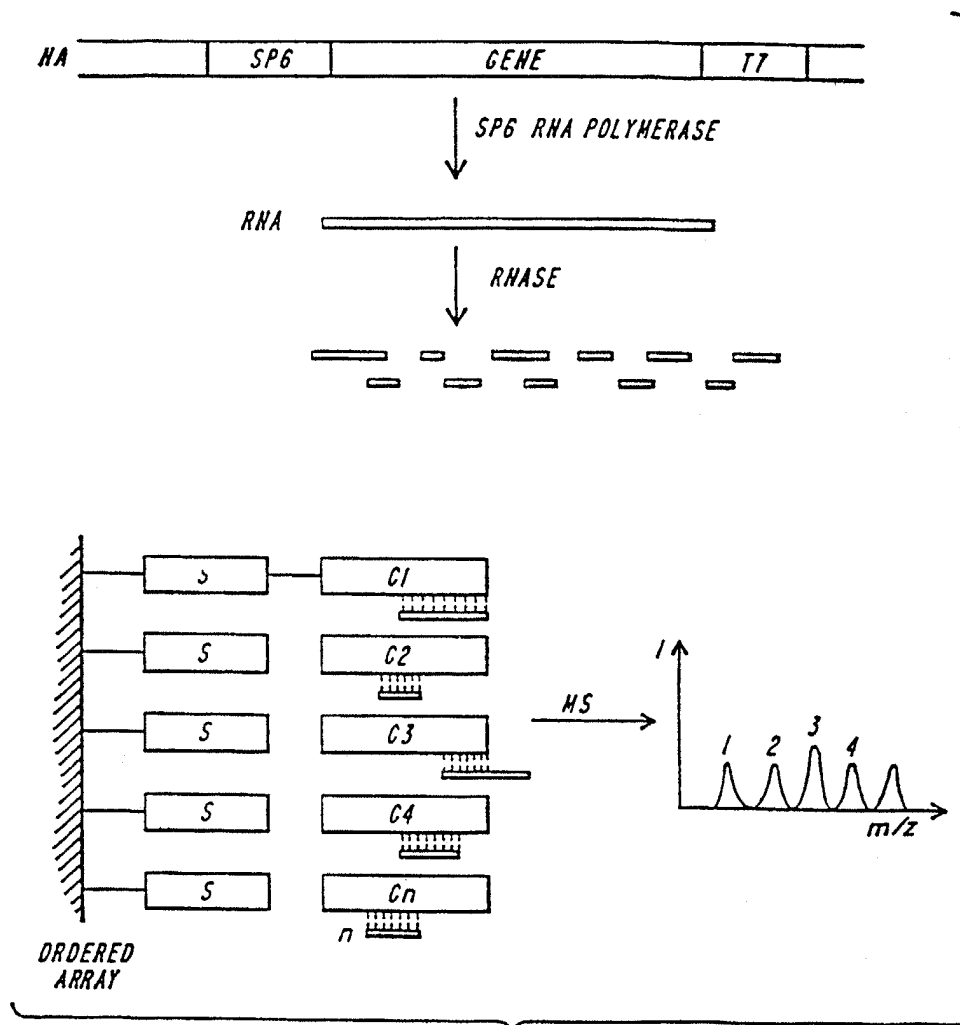
6/123

**FIG. 6A****FIG. 6B****FIG. 6C**



8/123

**FIG. 8**

**FIG. 9**

10/123

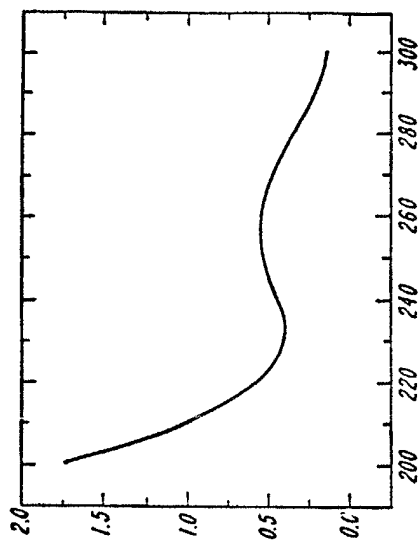


FIG. 10A-2

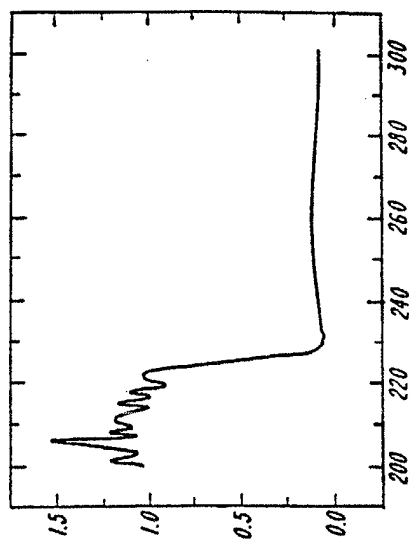


FIG. 10A-4

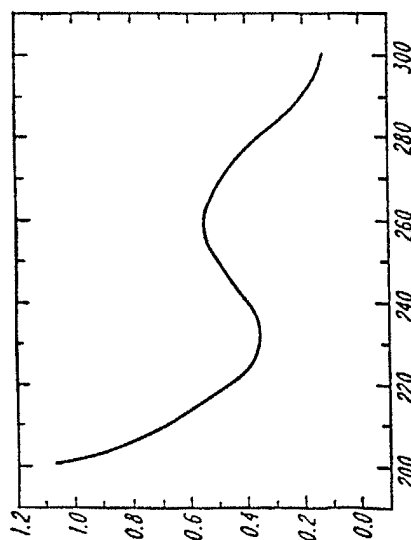


FIG. 10A-1

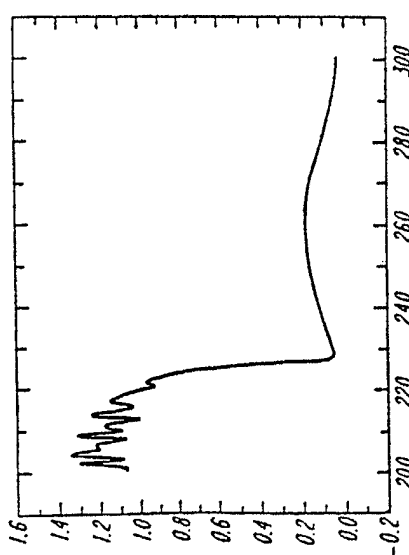
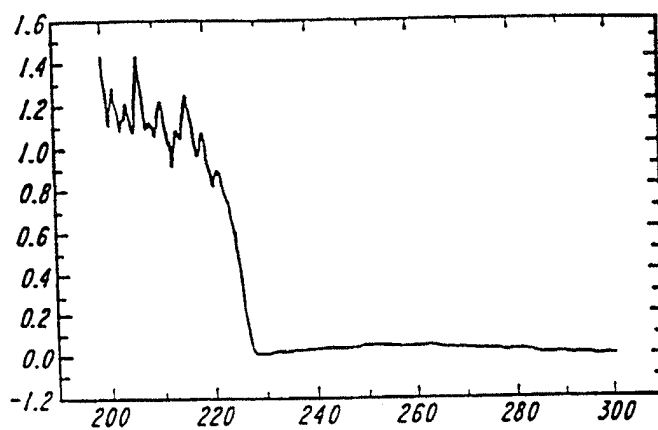
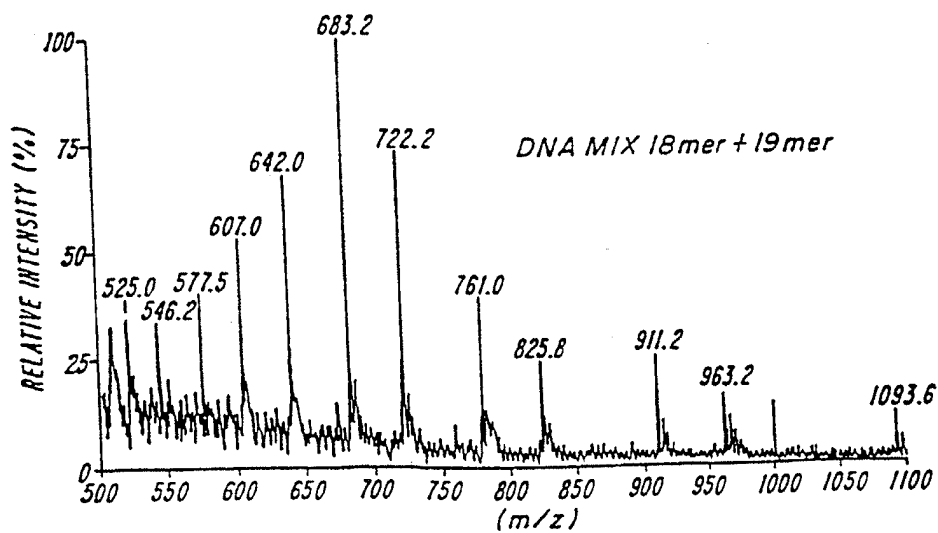


FIG. 10A-3

11 / 123

**FIG. 10B****FIG. 12A**

12/123

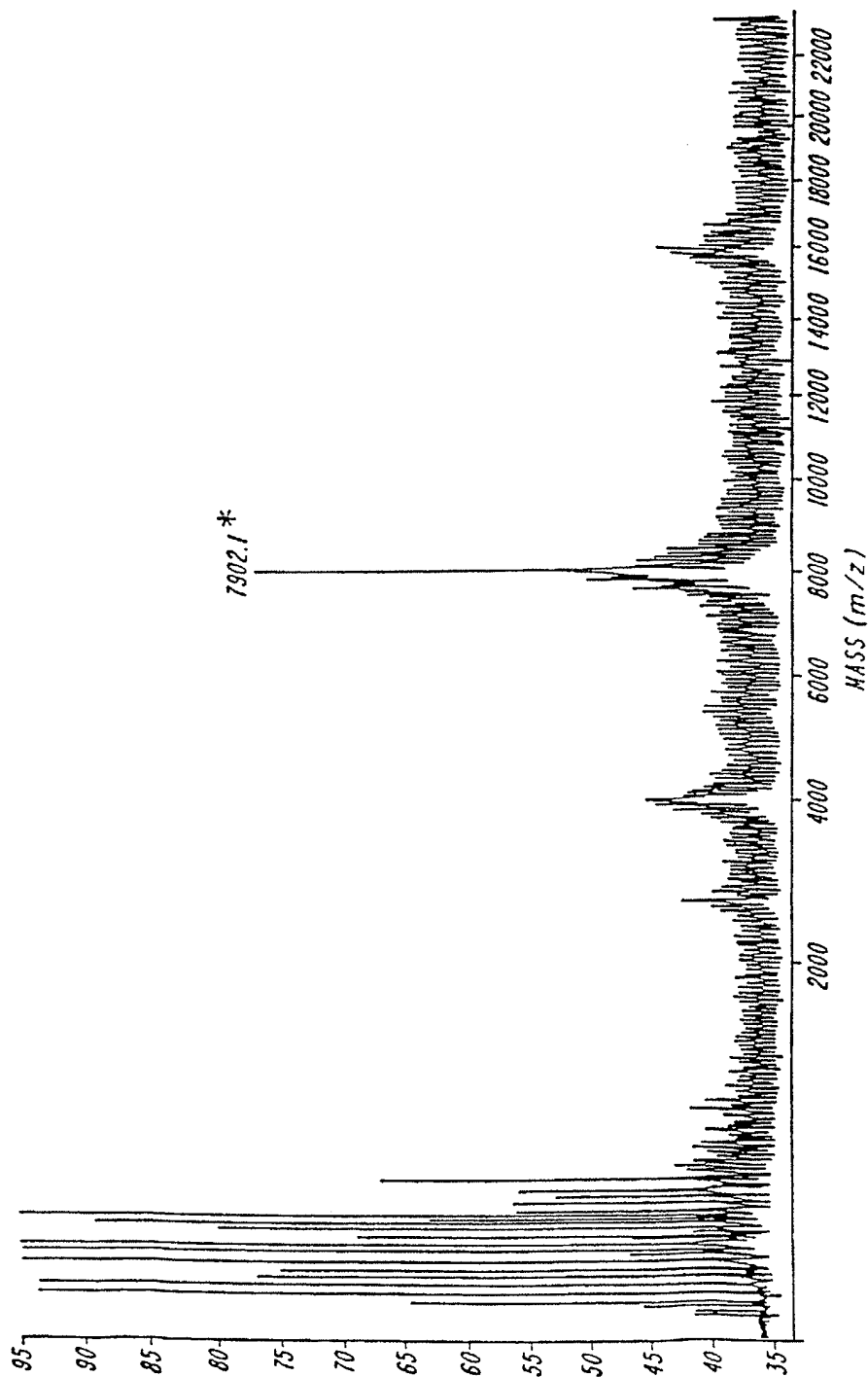


FIG. 10C

SUBSTITUTE SHEET (RULE 26)

13 / 123

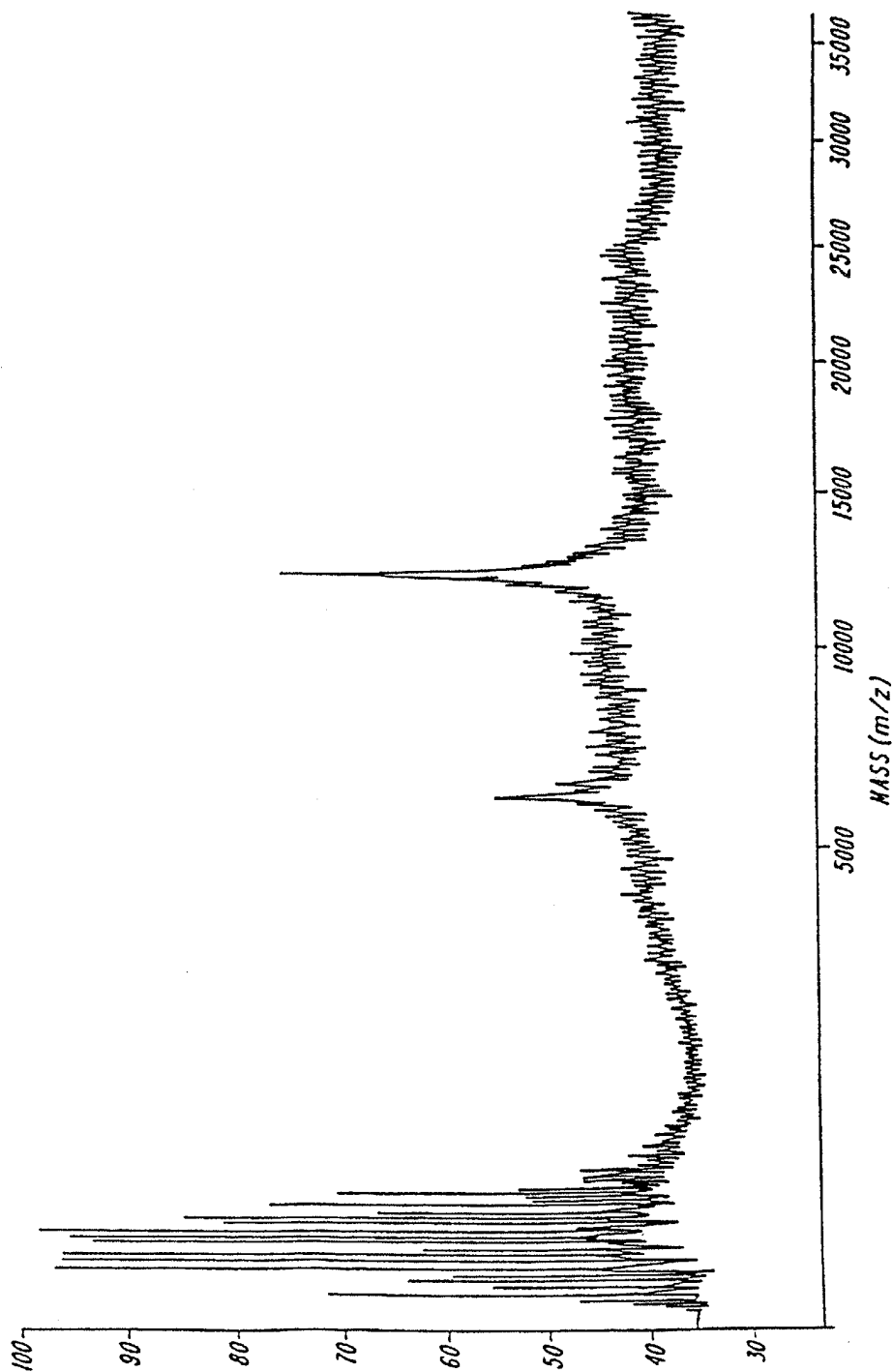


FIG. 11

SUBSTITUTE SHEET (RULE 26)

14 / 123

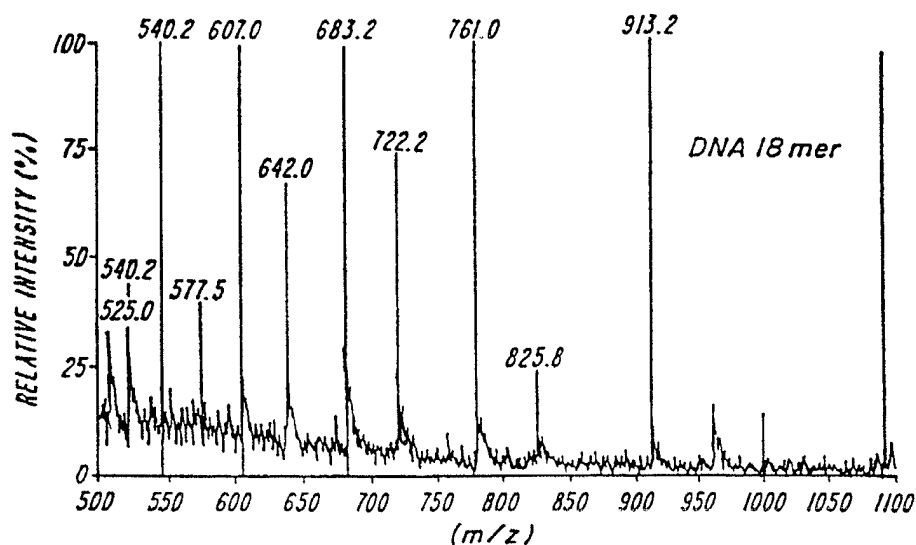


FIG. 12B

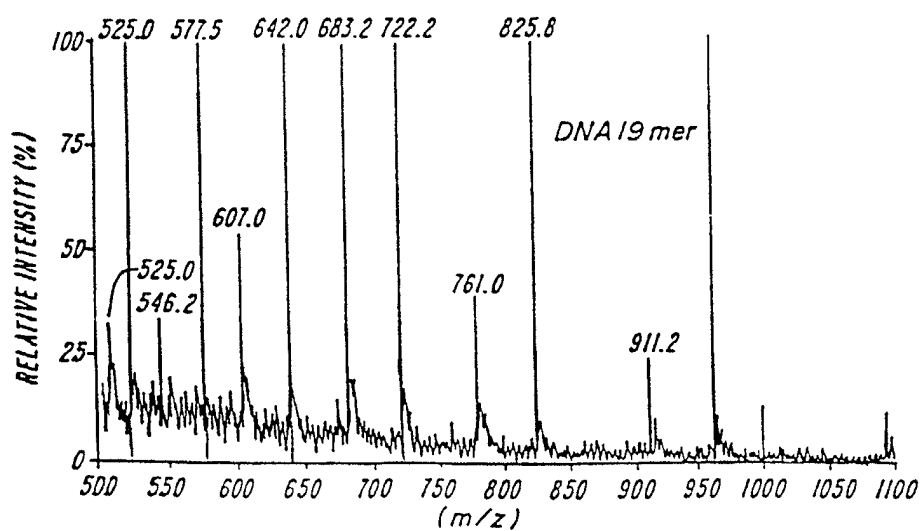


FIG. 12C

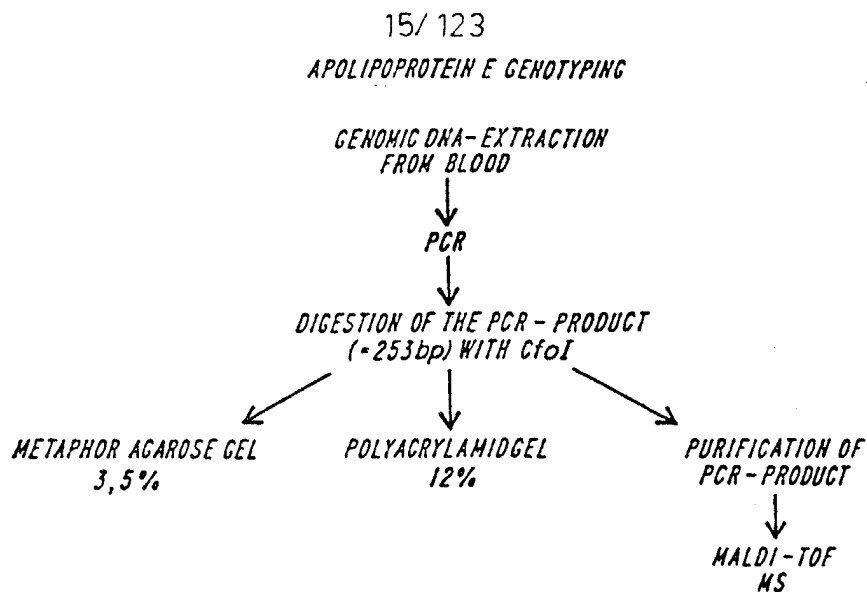


FIG. 19

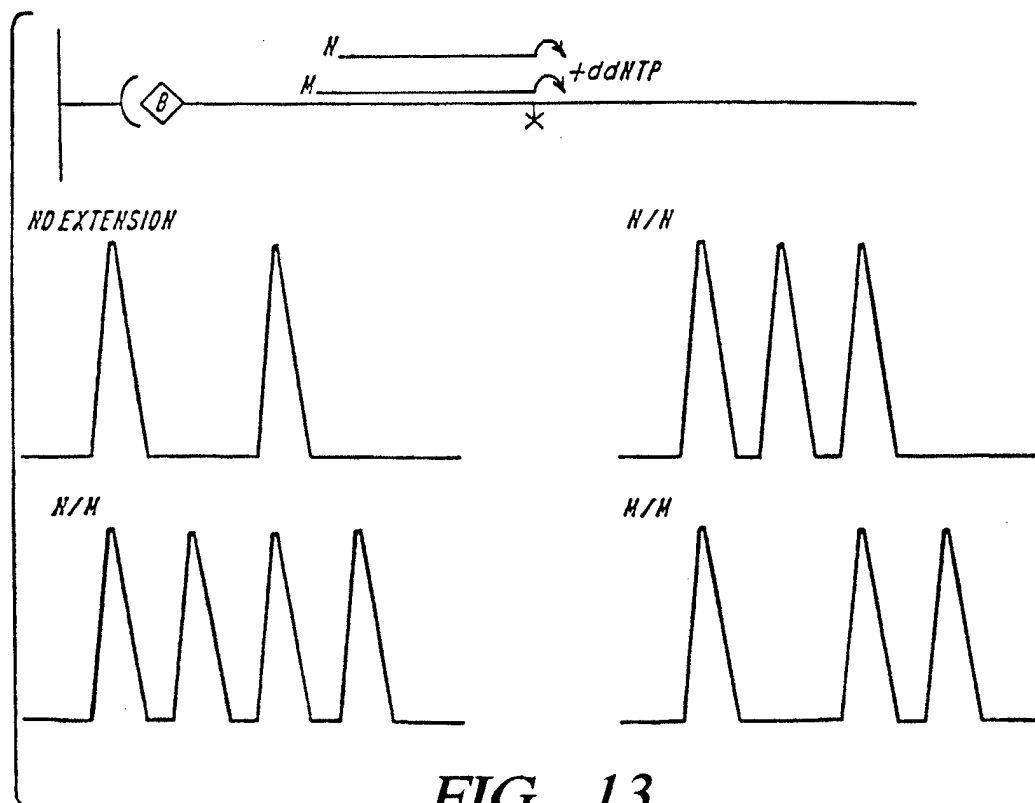


FIG. 13

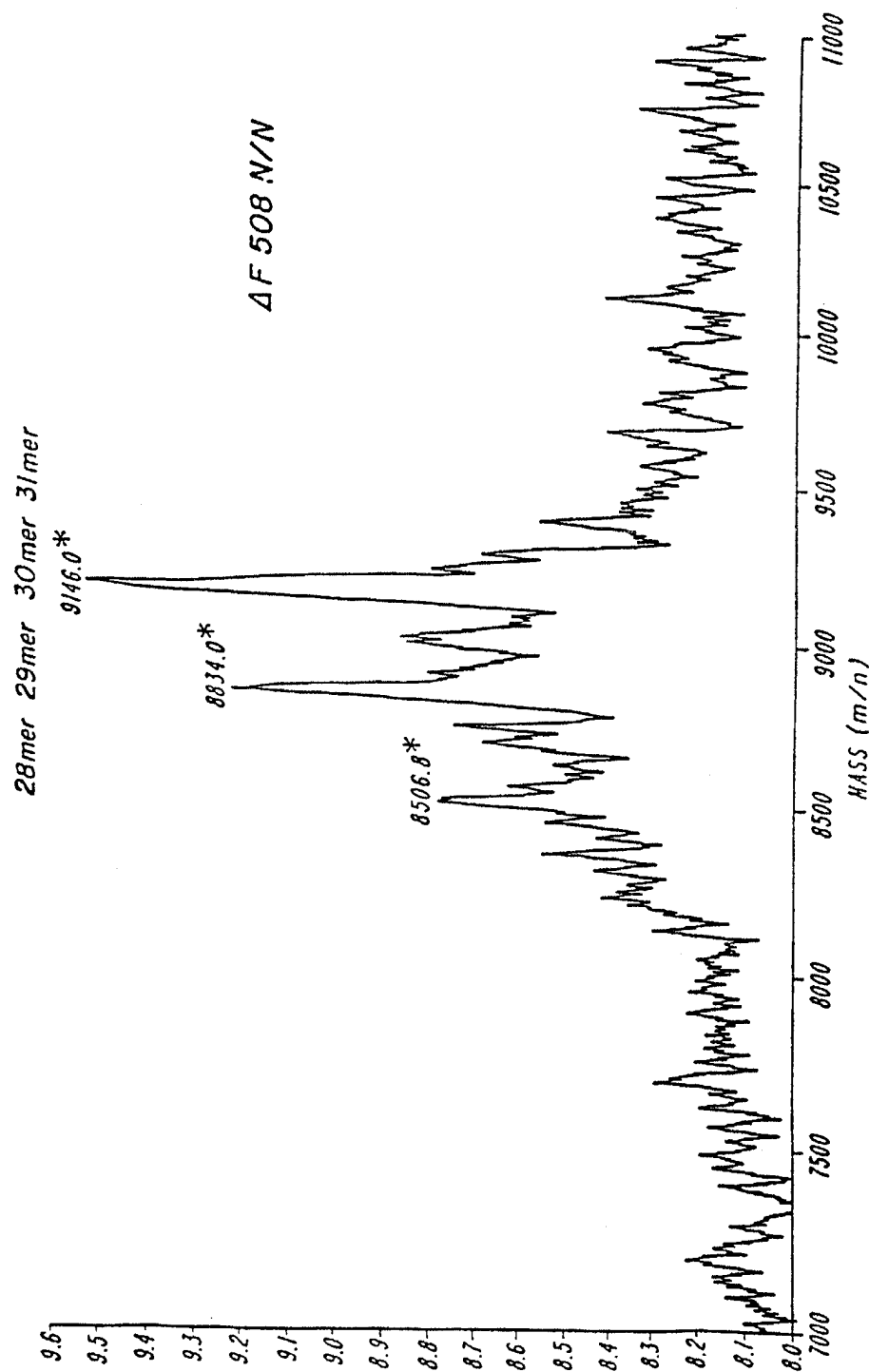


FIG. 14

17/123

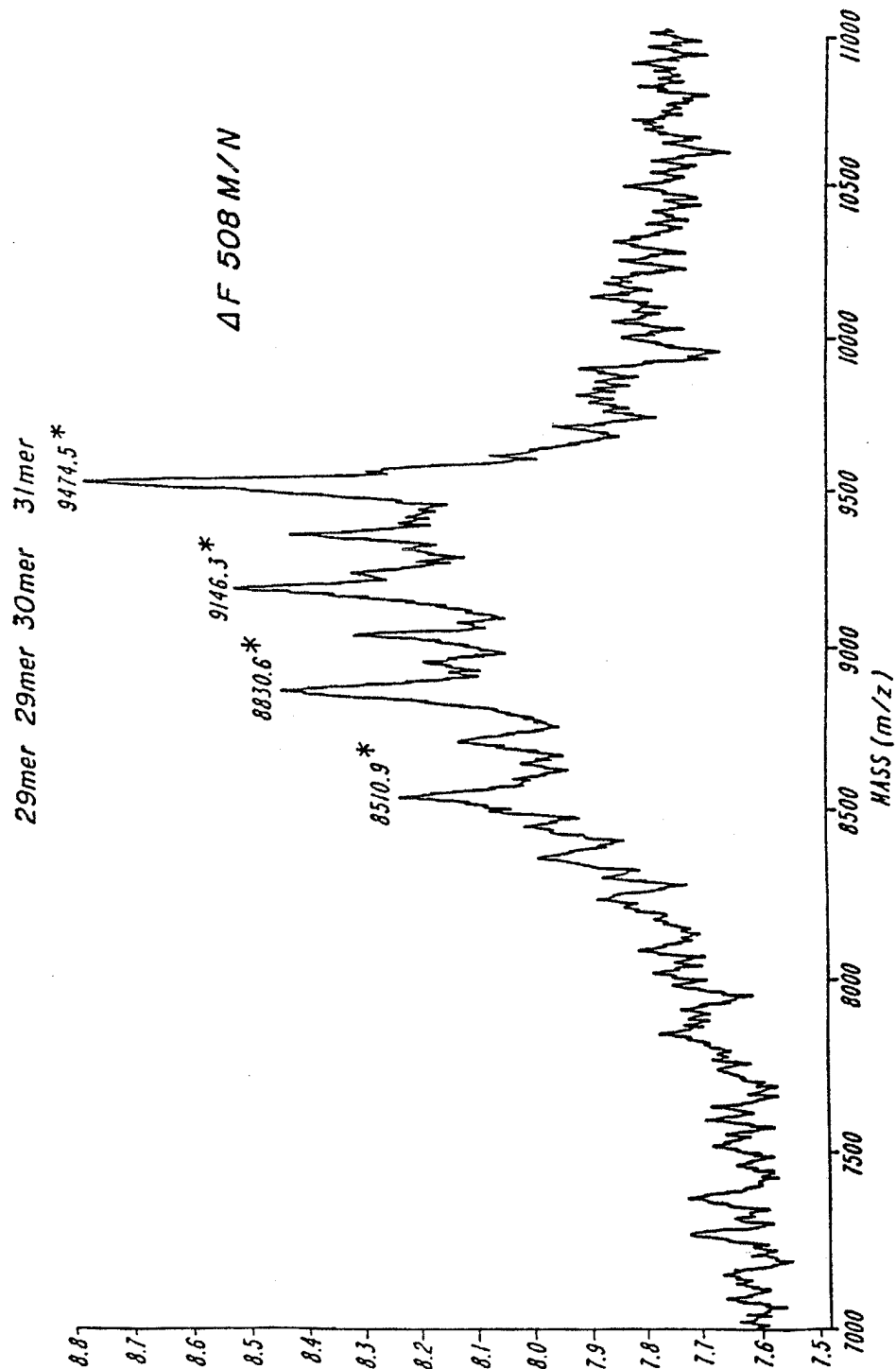


FIG. 15

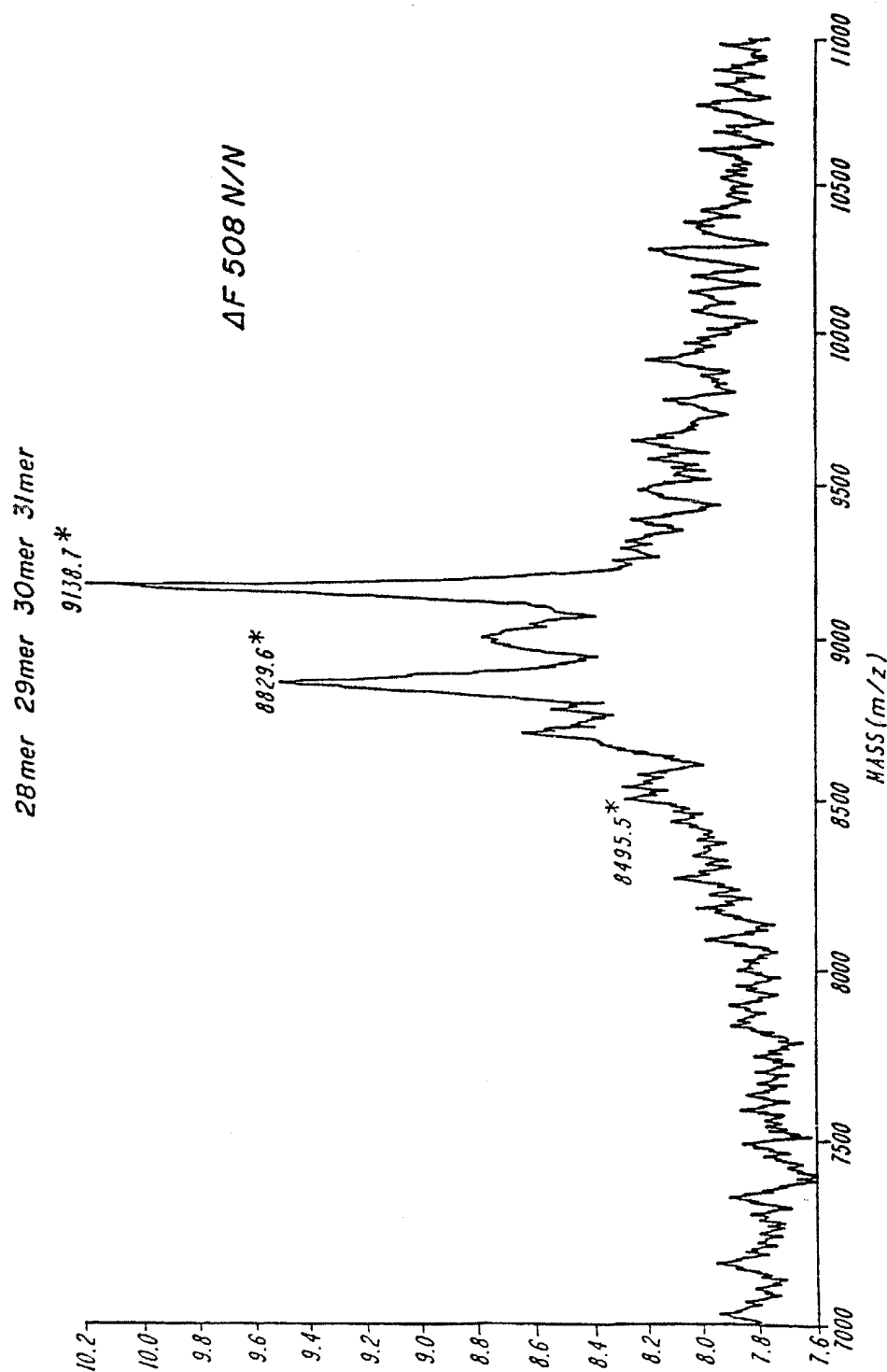


FIG. 16

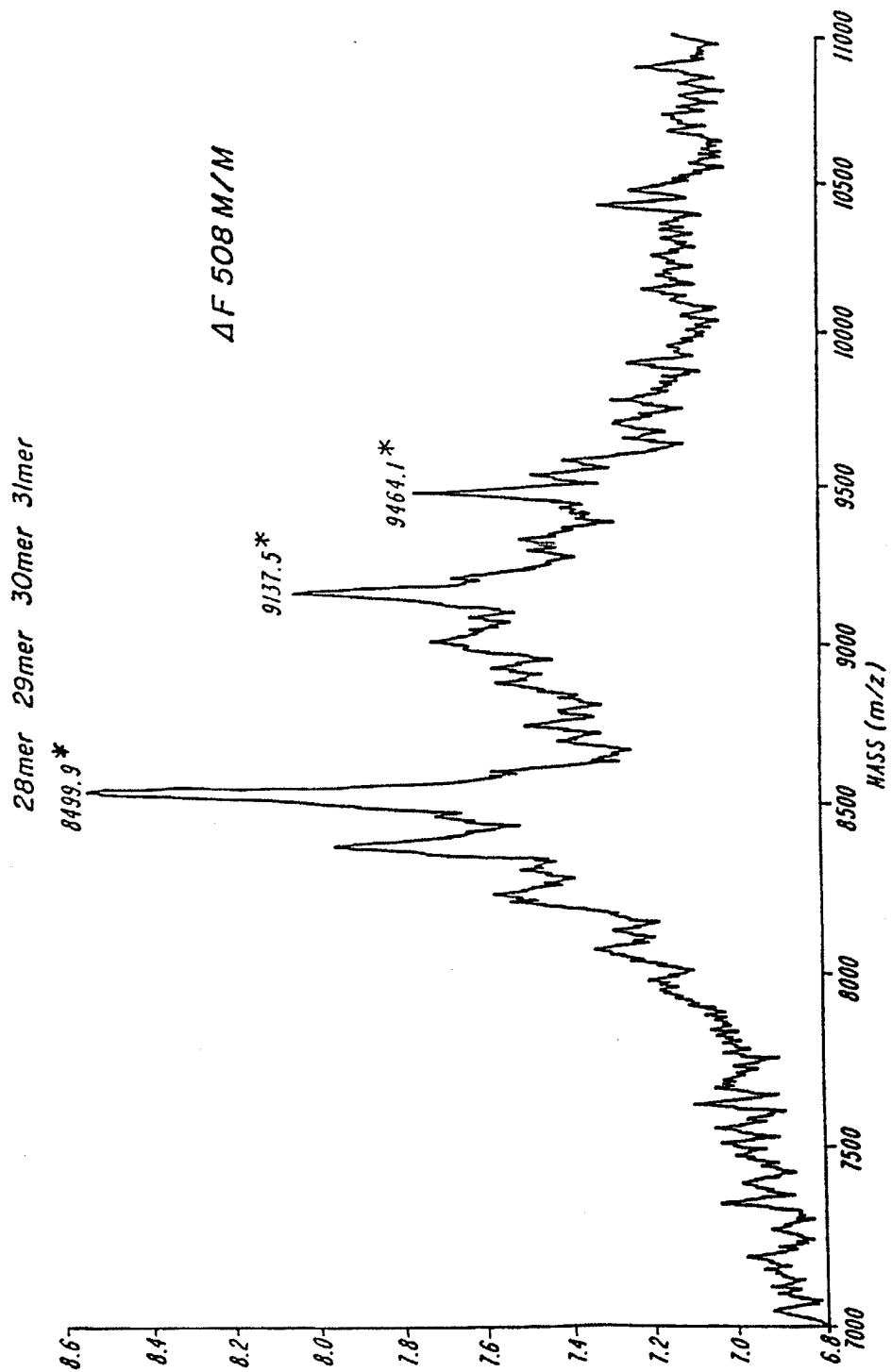


FIG. 17

20/123

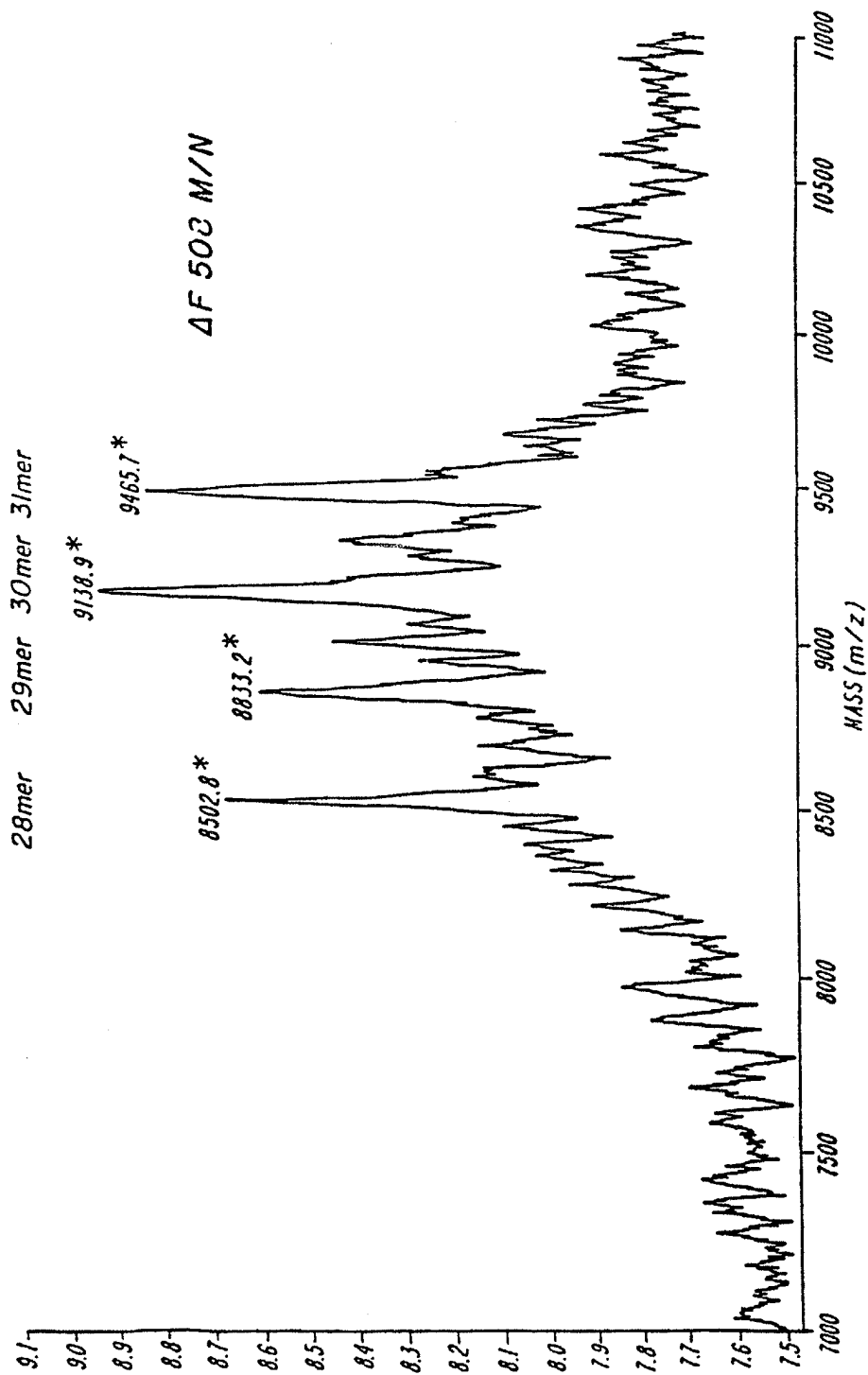


FIG. 18



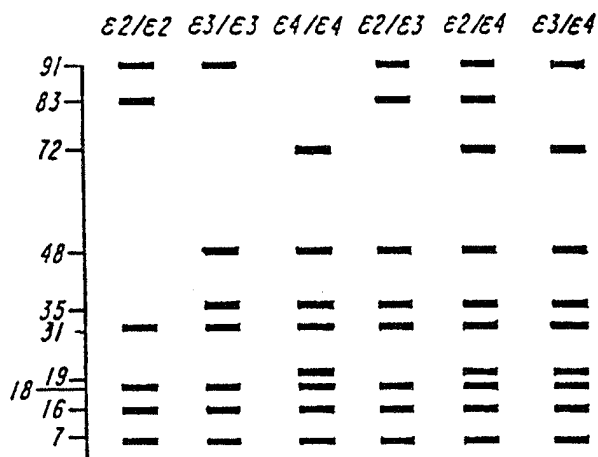


FIG. 21A

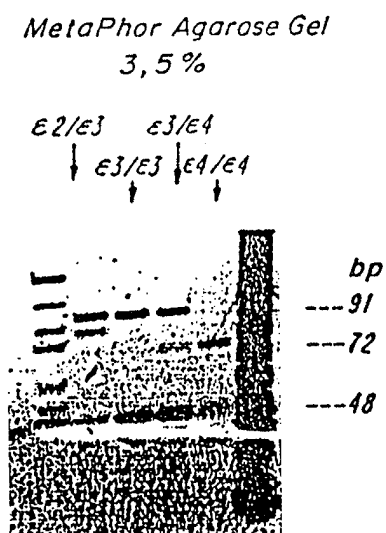


FIG. 21B

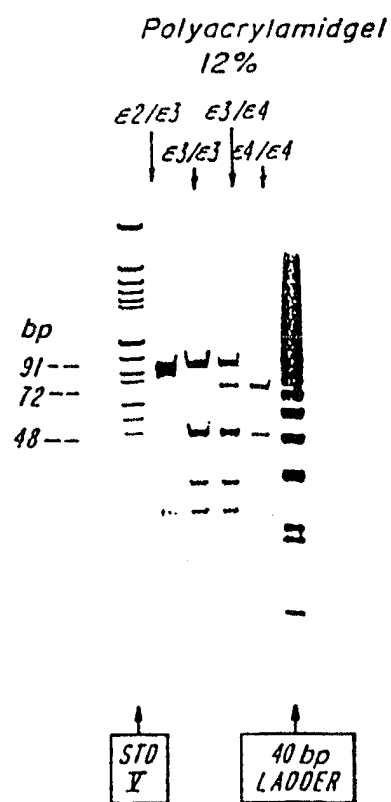


FIG. 21C

24/123

MOLECULAR WEIGHT OF THE VARIABLE FRAGMENTS IN Da:

			$\epsilon 2/\epsilon 2$	$\epsilon 3/\epsilon 3$	$\epsilon 4/\epsilon 4$	$\epsilon 2/\epsilon 3$	$\epsilon 2/\epsilon 4$	$\epsilon 3/\epsilon 4$
91 bp	SENSE	28421	X	X		X	X	X
	ANTISENSE	27864						
83bp	SENSE	25747	X			X	X	
	ANTISENSE	25591						
72bp	SENSE	22440			X		X	X
	ANTISENSE	21494						
48bp	SENSE	14844		X	X	X	X	X
	ANTISENSE	14857						
35bp	SENSE	10921		X	X	X	X	X
	ANTISENSE	10751						

FIG. 22A

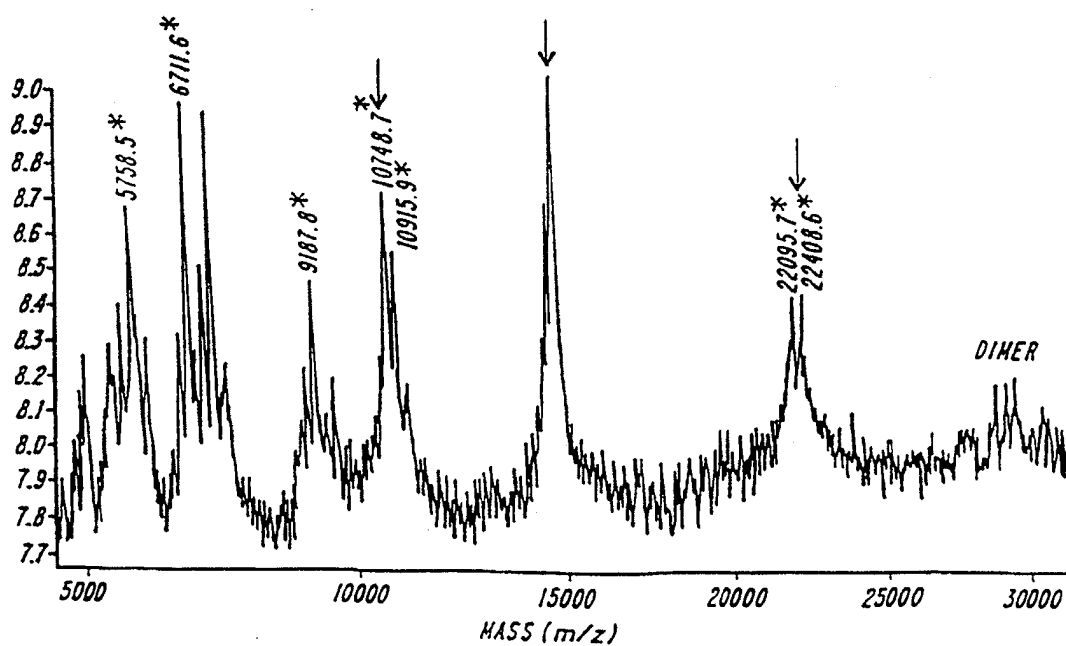


FIG. 22B

25/123

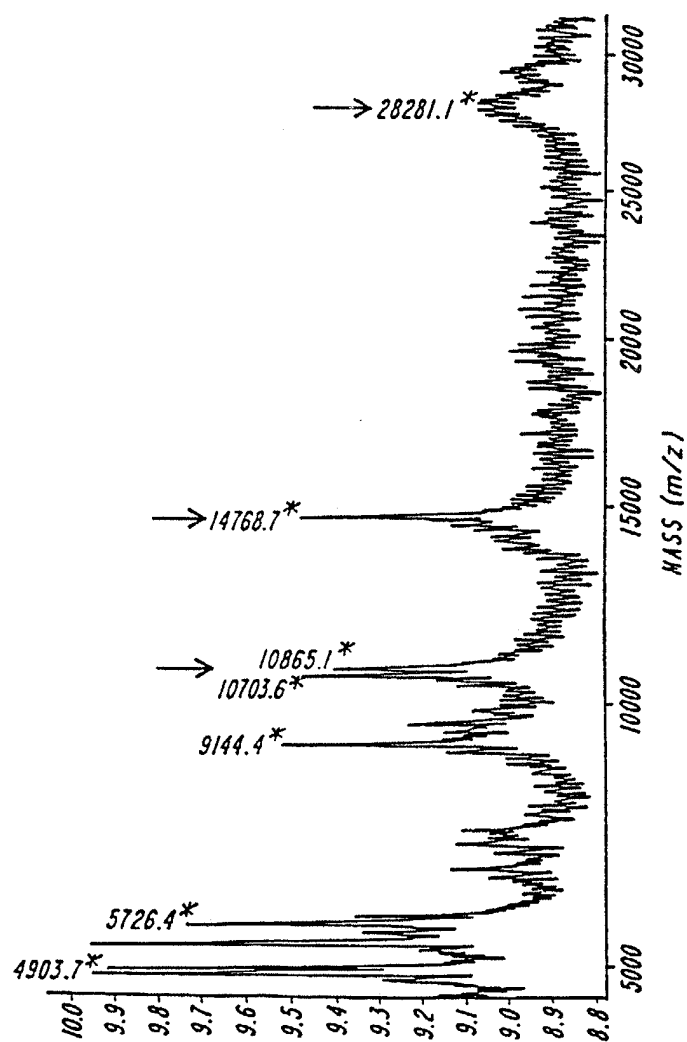


FIG. 23A

26 / 123

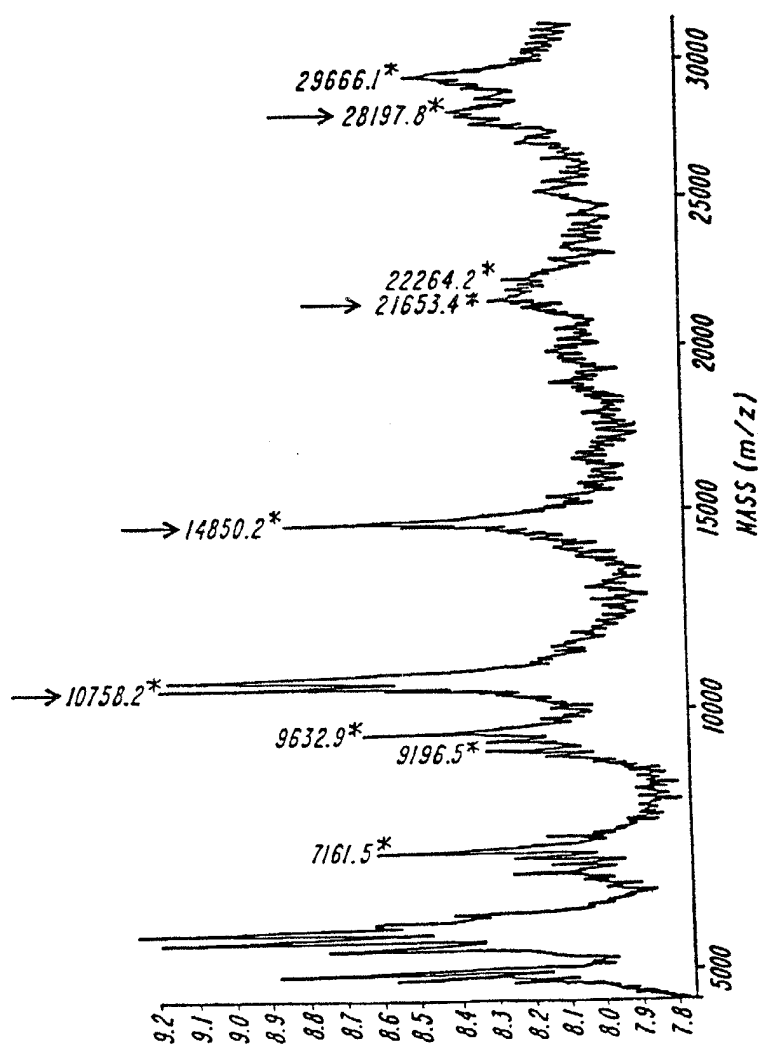


FIG. 23B



M 1 2 3 4 5 6

FIG. 24

28 / 123

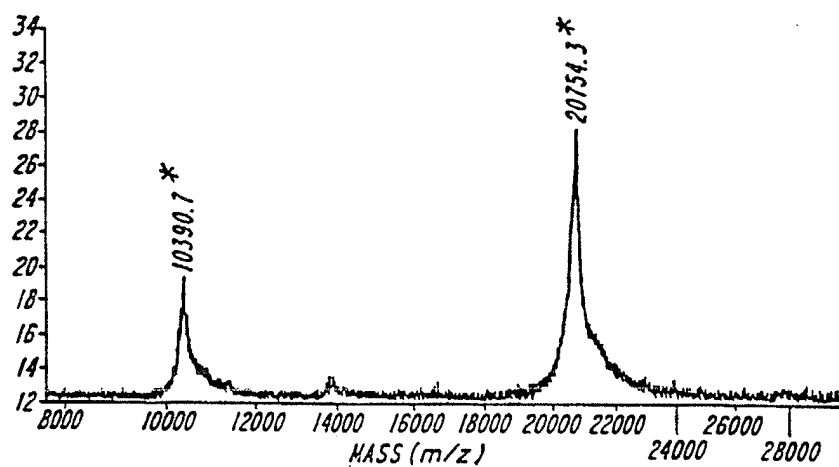


FIG. 25A

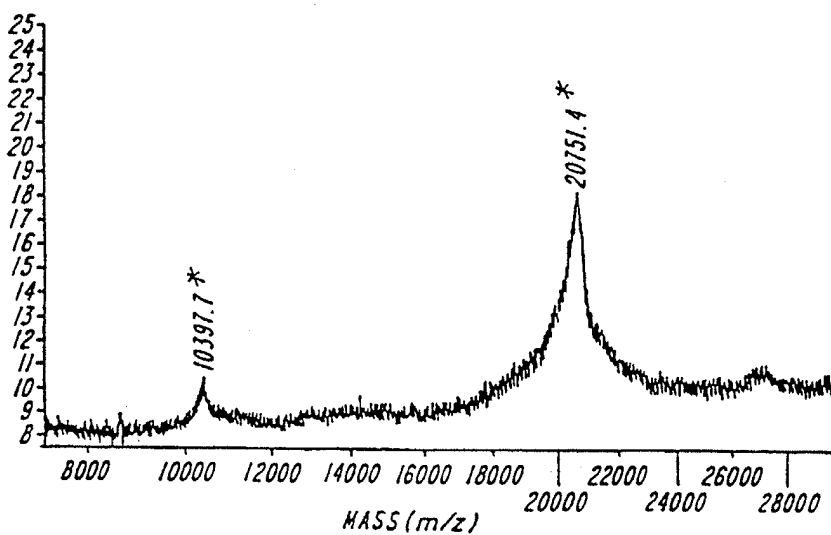


FIG. 25B

29 / 123

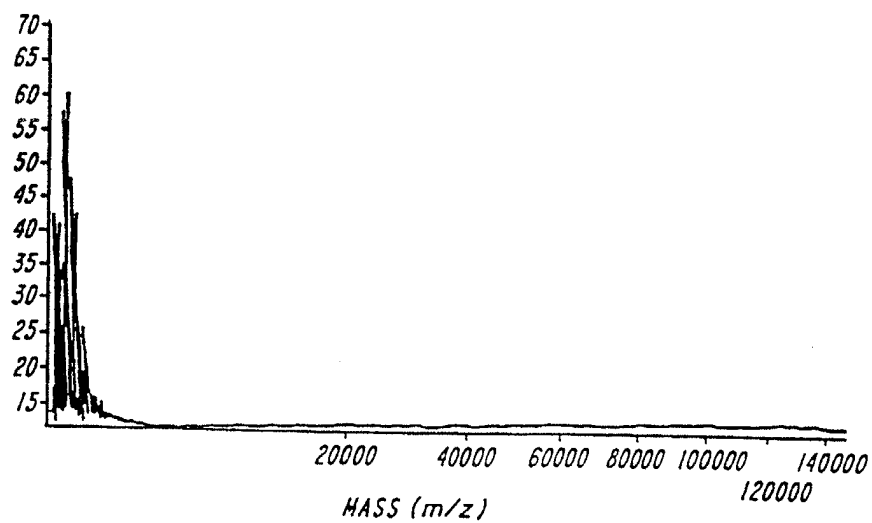


FIG. 25C

30/123

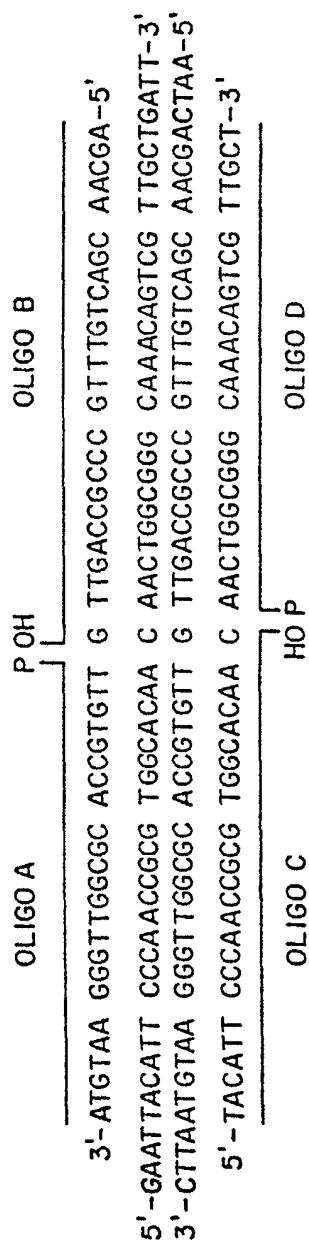


FIG. 26

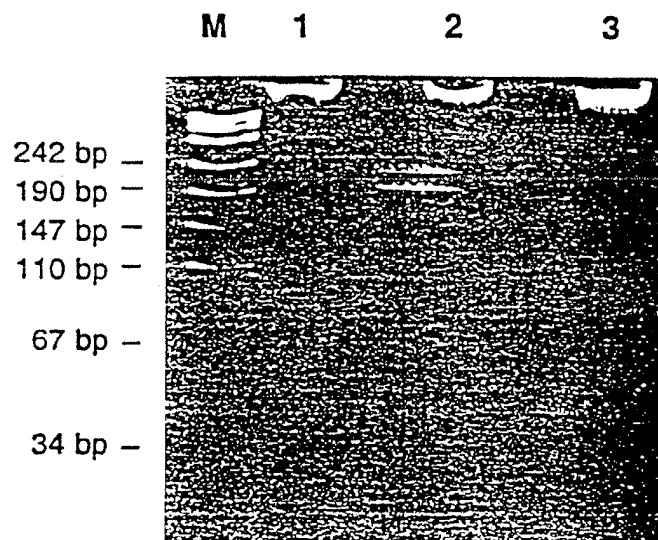
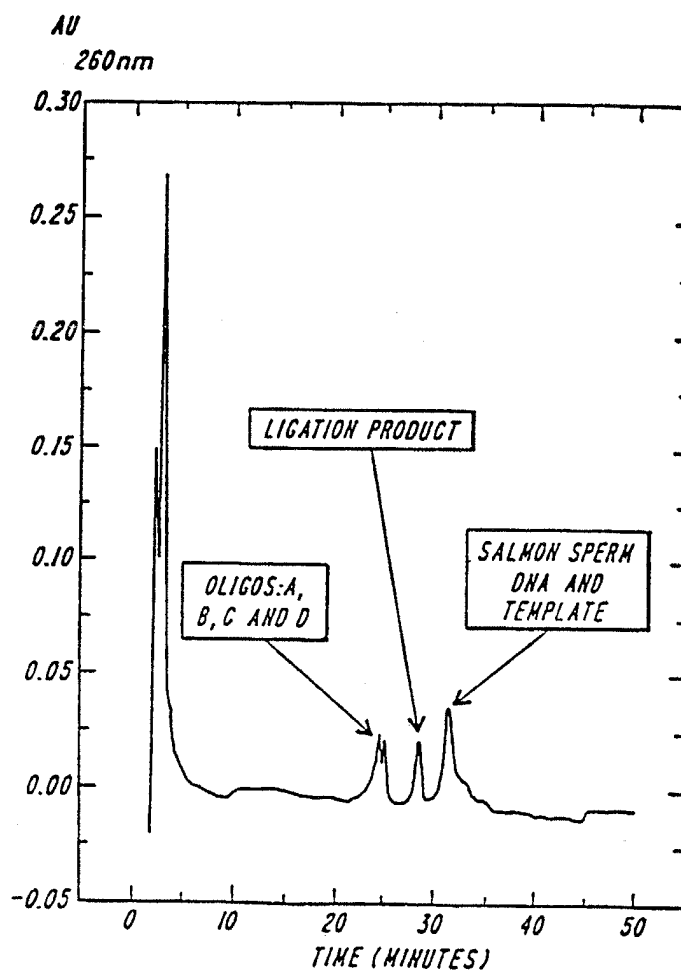
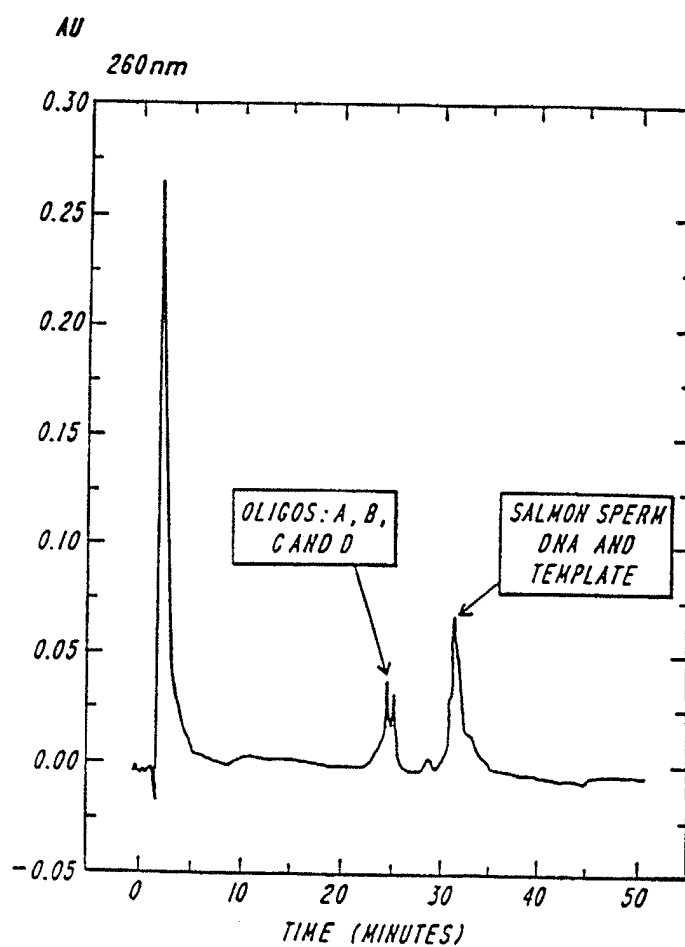


FIG. 27

32/123

**FIG. 28**

**FIG. 29**

34 / 123

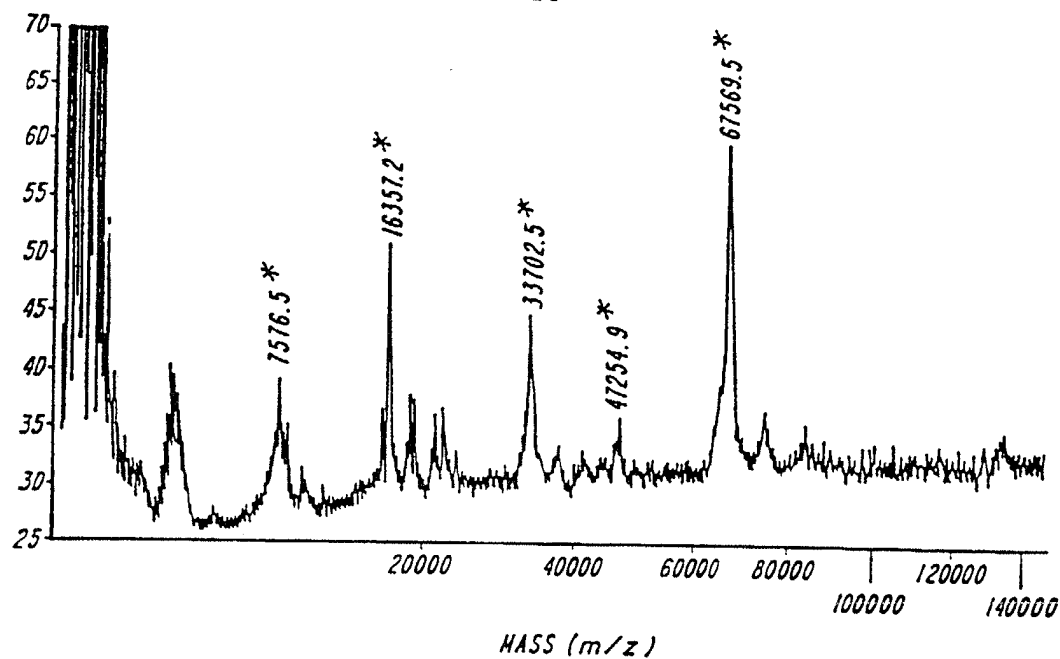


FIG. 30A

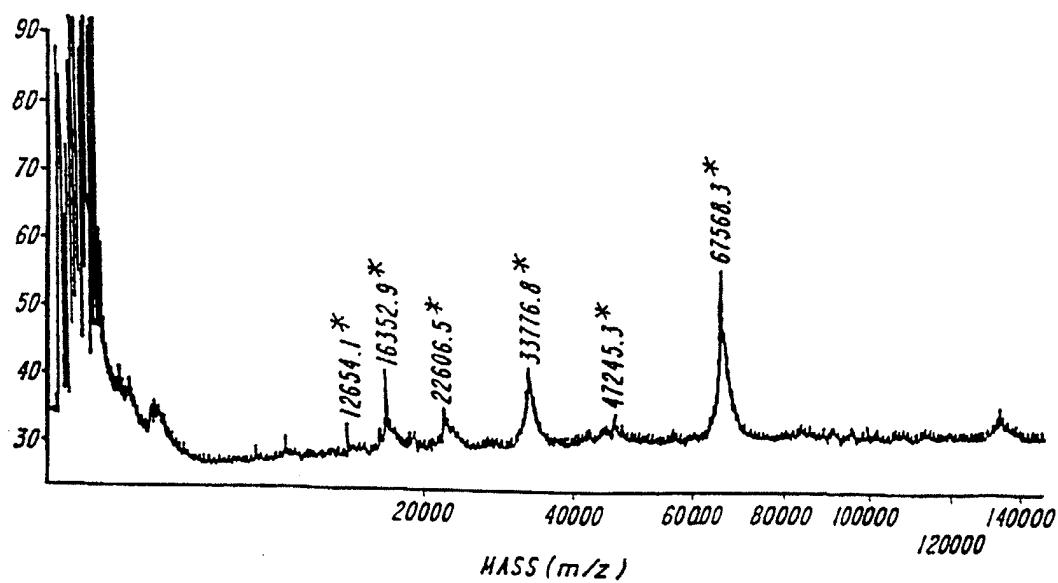
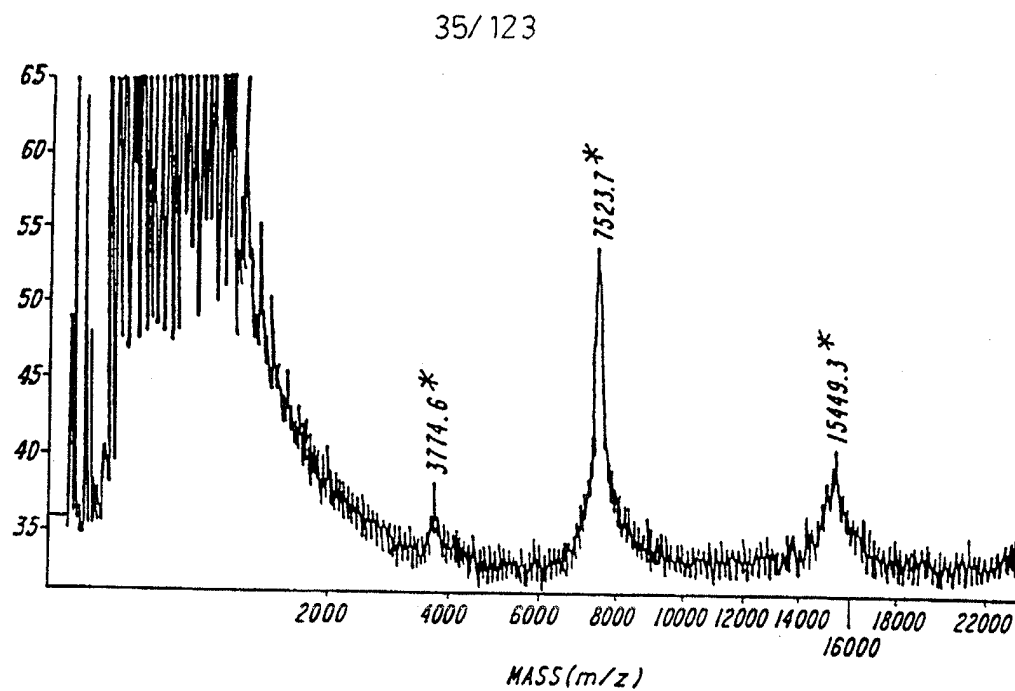
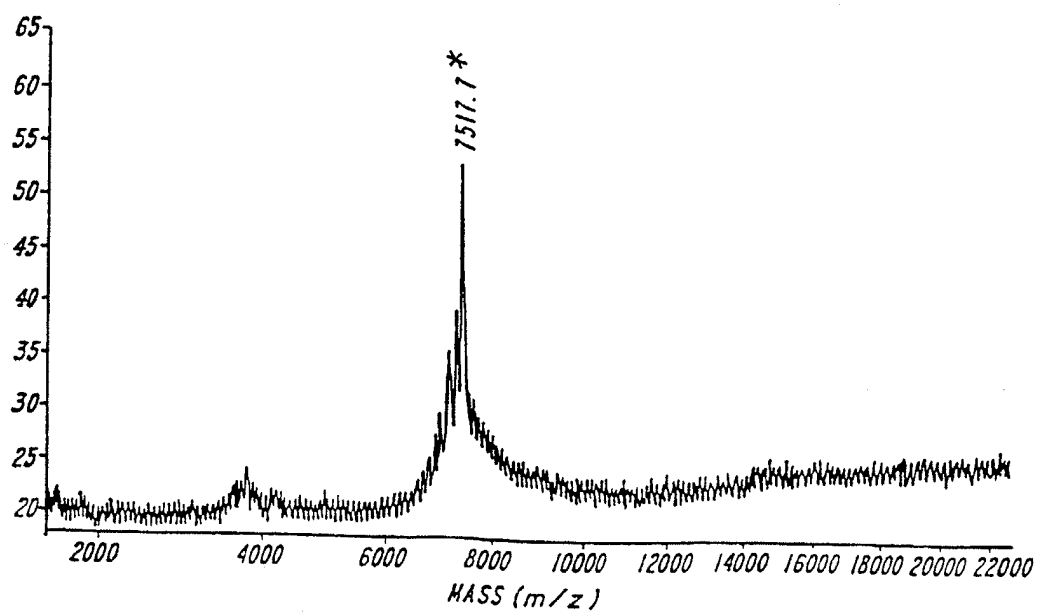
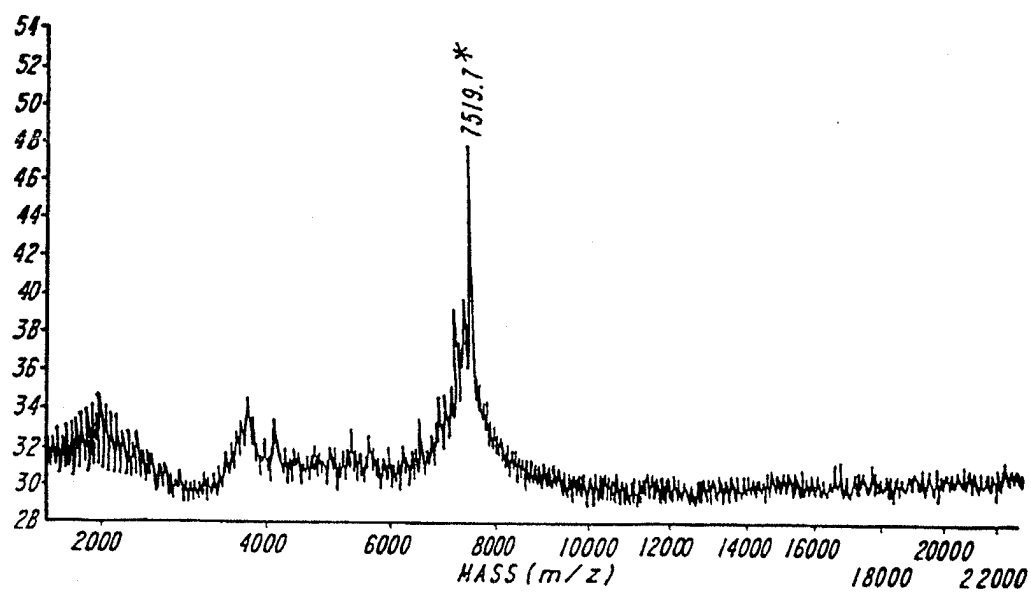


FIG. 30B

*FIG. 31A**FIG. 31B*

36 / 123

*FIG. 32*

37 / 123

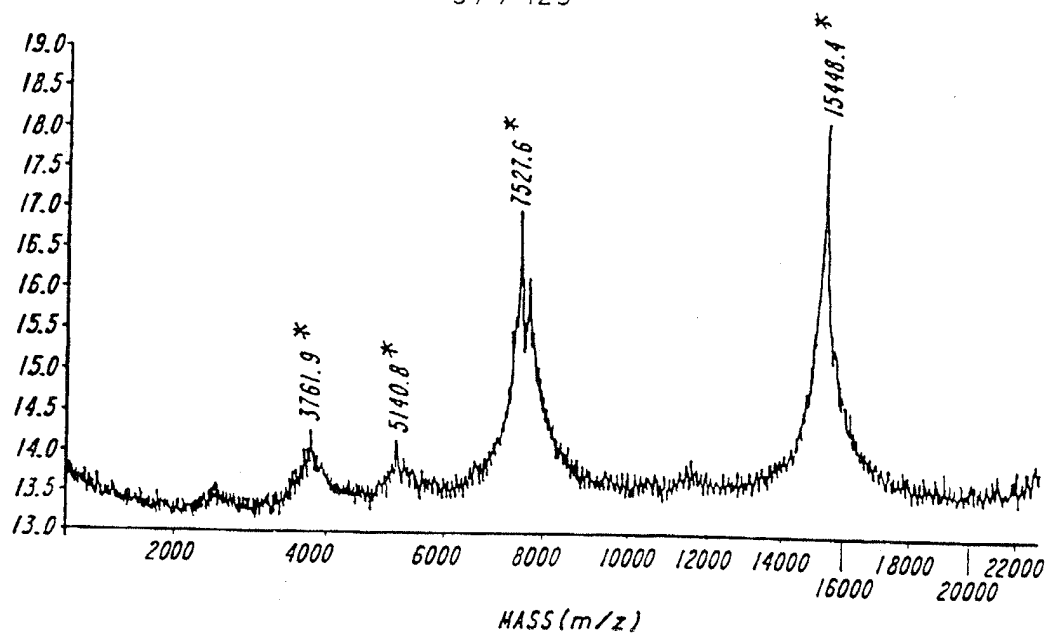


FIG. 33A

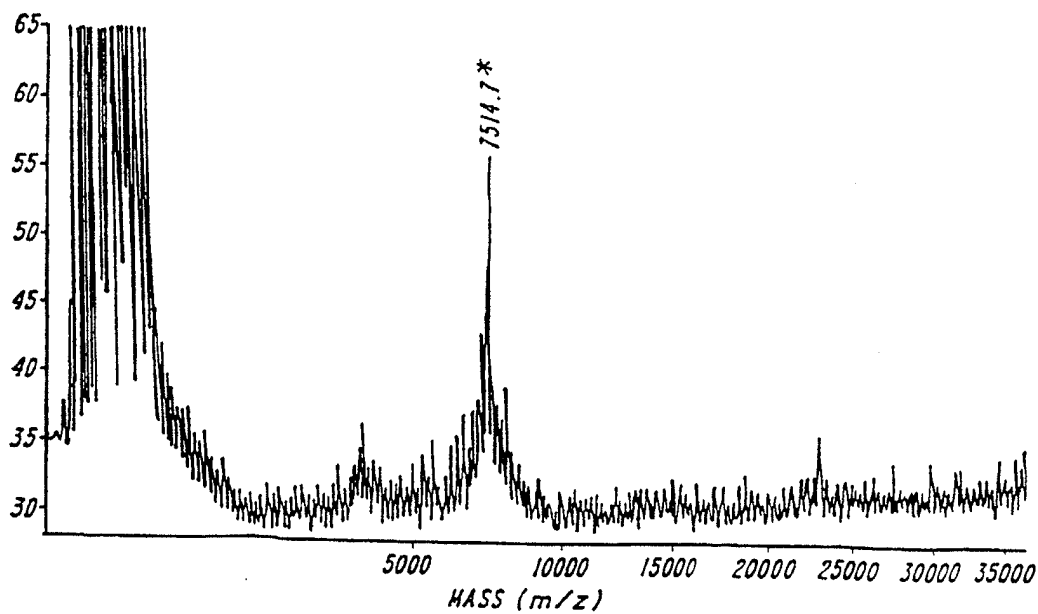


FIG. 33B

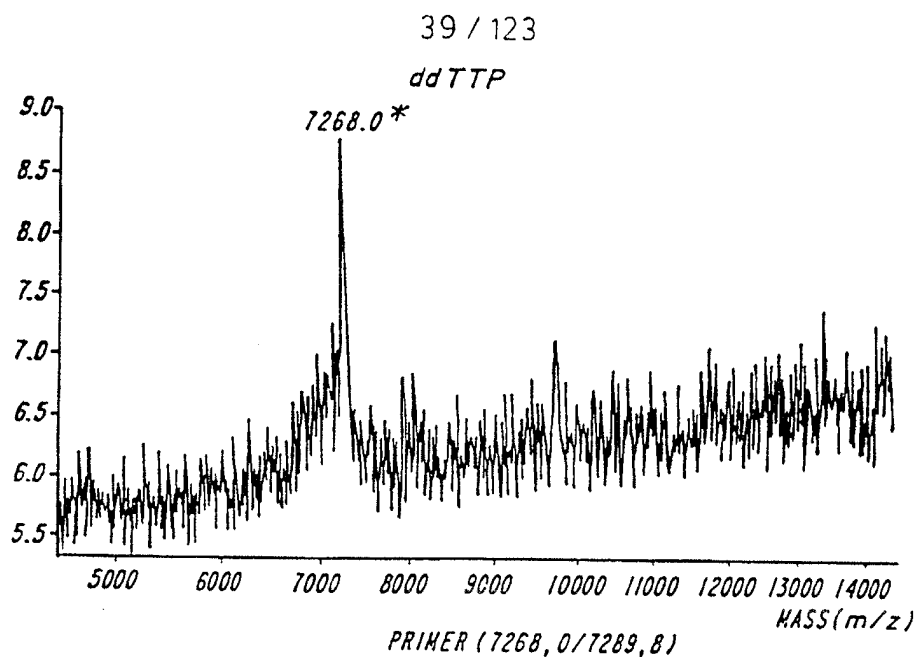
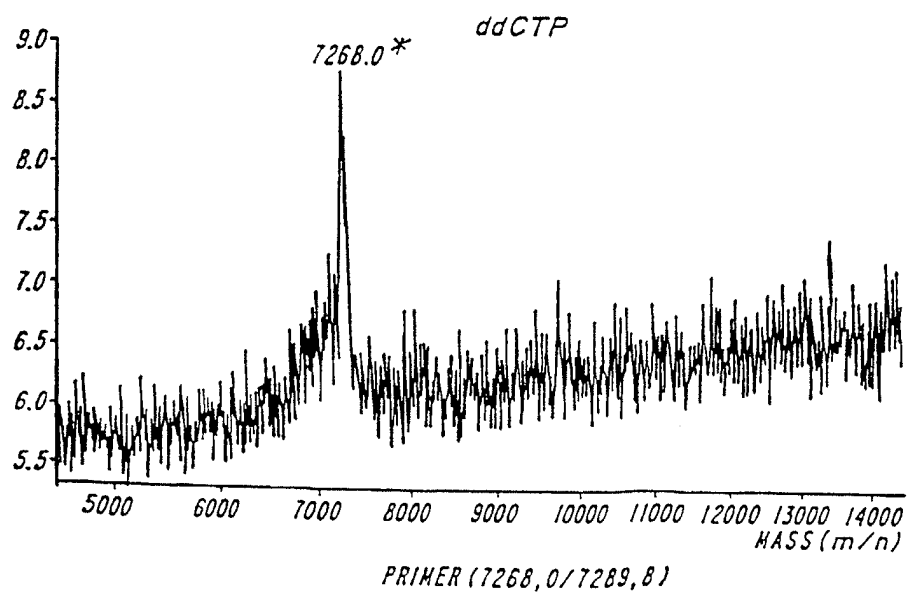
38/123

506507508
IleIlePhe
ACCATTAAAGAAATATCATCTTTGGTGTTCCTATGATGAATATAGAAGCGTCATC
primer ACCACAAAGGATACTACTTTATATC (7289,8)
wildtype TAGAAACCCACAAAGGATACTACTTTATATC (8846,8)
ΔF508 TA---ACCACAAAGGATACTACTTTATATC (7891,2)
ΔI507 TAG---AAACCCACAAAGGATACTACTTTATATC (8846,8)

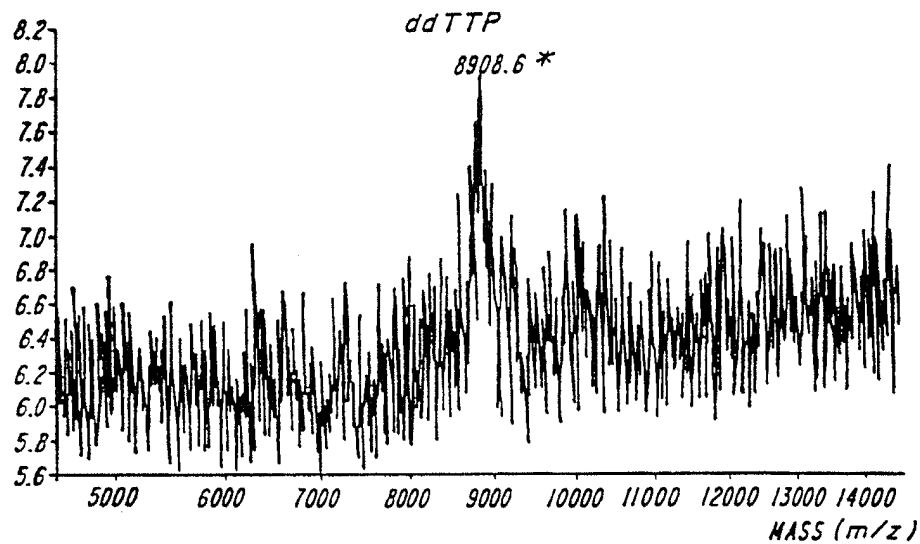
FIG. 34A

506507508
IleIlePhe
ACCATTAAAGAAATATCATCTTTGGTGTTCCTATGATGAATATAGAAGCGTCATC
primer ACCACAAAGGATACTACTTTATATC (7289,8)
wildtype CTTTATATAGTAGAACCCACAAAGGATACTACTTTATATC (11612,6)
ΔF508 CTTTATATAGTA---ACCACAAAGGATACTACTTTATATC (10657,0)
ΔI507 CTTTATATAG---AAACCCACAAAGGATACTACTTTATATC (10666,0)
506Ser CGTAGAAACCCACAAAGGATACTACTTTATATC (9465,2)

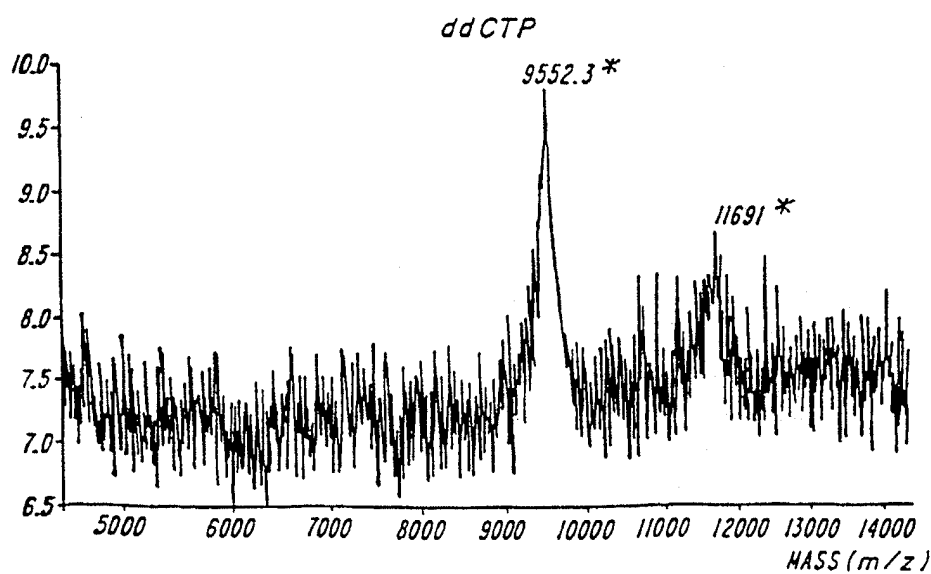
FIG. 34B

**FIG. 35A****FIG. 35B**

40/123

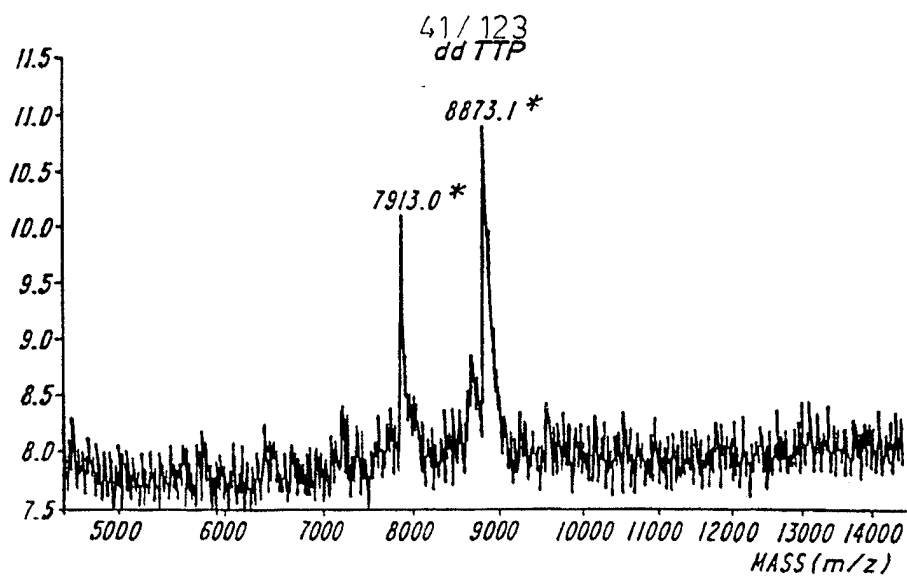


WILDTYPE (8908, 6 / 8846, 8)

FIG. 35C

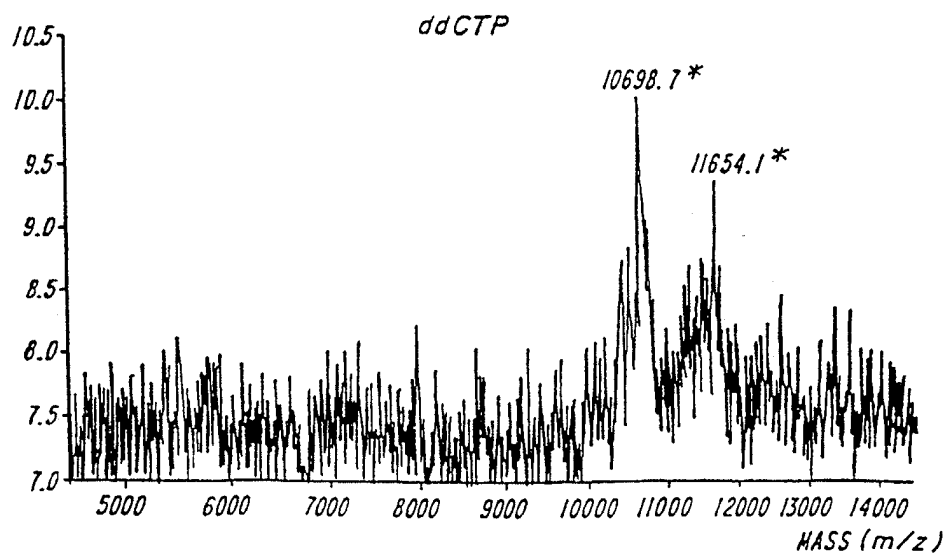
506S (9552, 3 / 9465, 2)
WILDTYPE (11691, 9 / 11612, 6)

FIG. 35D



$\Delta F508$ (7913,0/7891,2)
WILDTYPE (8873,1/8846,8)

FIG. 35E



$\Delta 508$ (10698,7/10657,0)
WILDTYPE (11654,1/11612,6)

FIG. 35F

SUBSTITUTE SHEET (RULE 26)

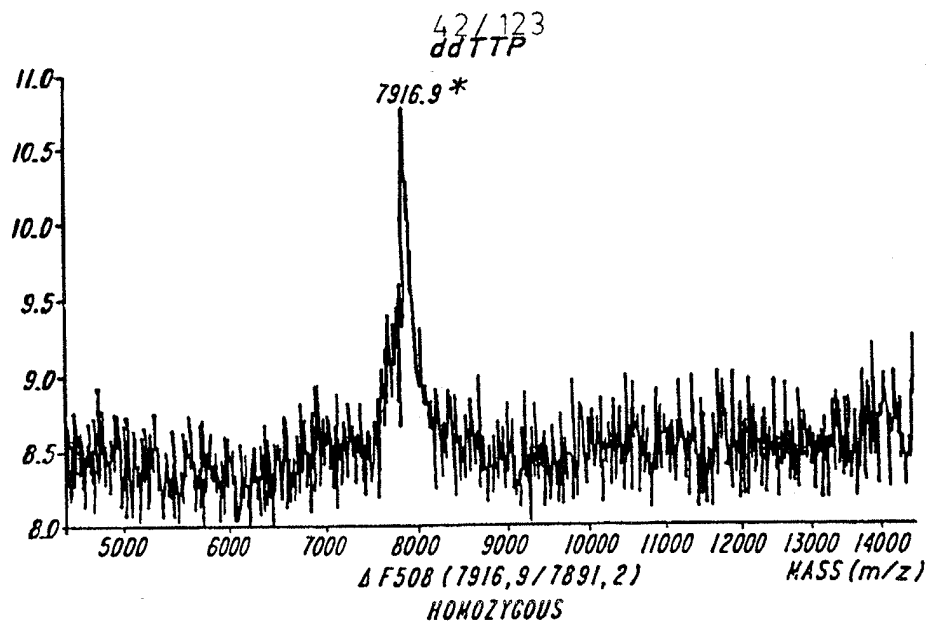


FIG. 35G

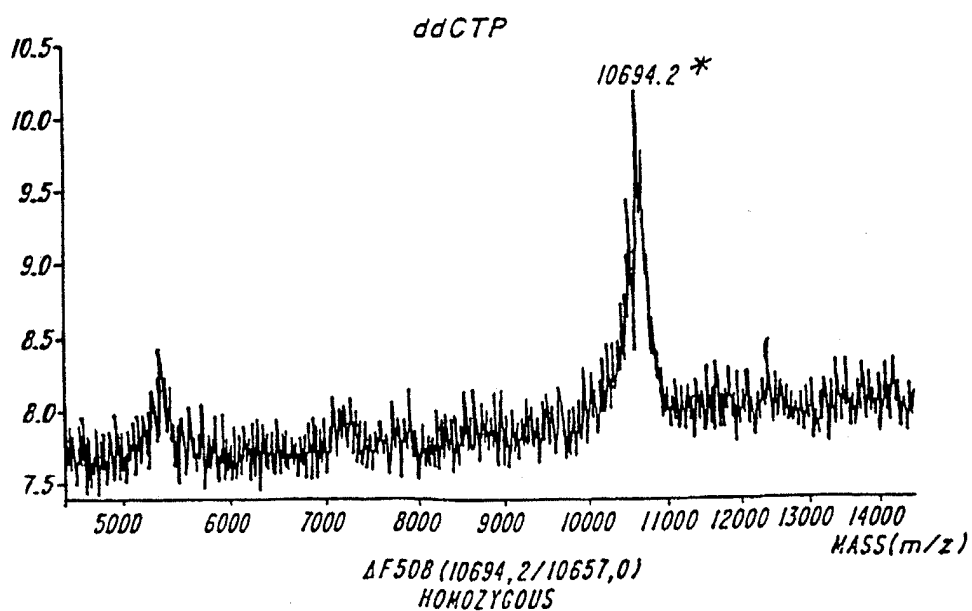


FIG. 35H

SUBSTITUTE SHEET (RULE 26)

43 / 123

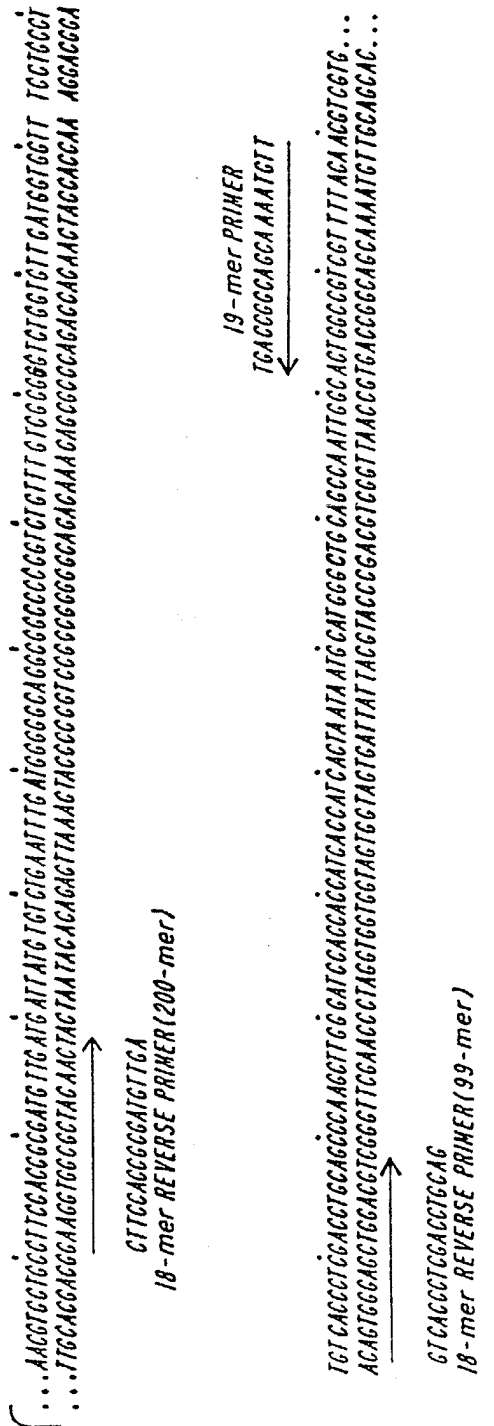


FIG. 36

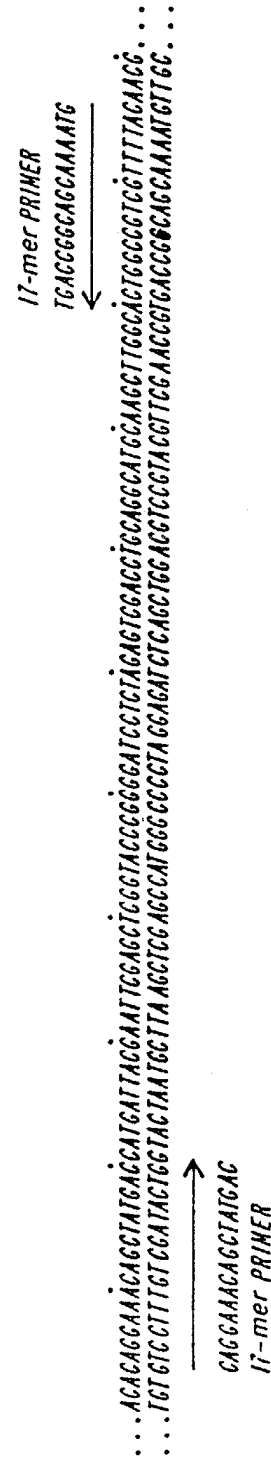


FIG. 37

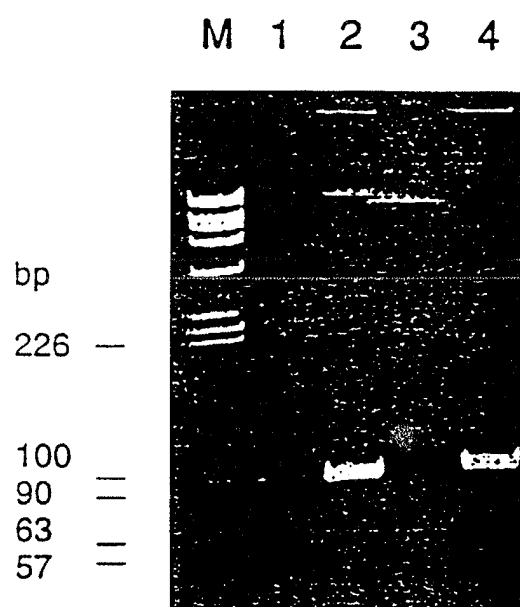


FIG. 38

1 2 3 4 5 6

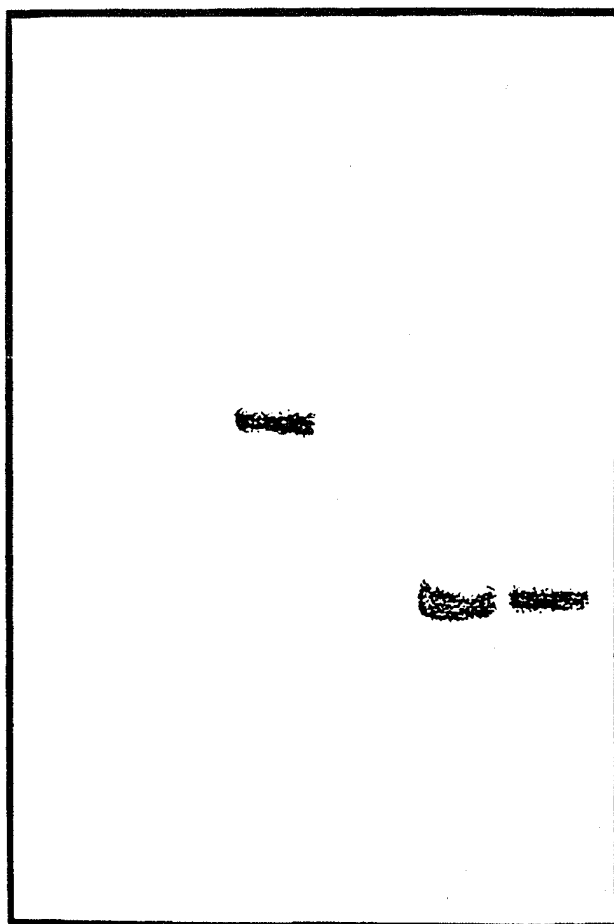


FIG. 39

46 / 123

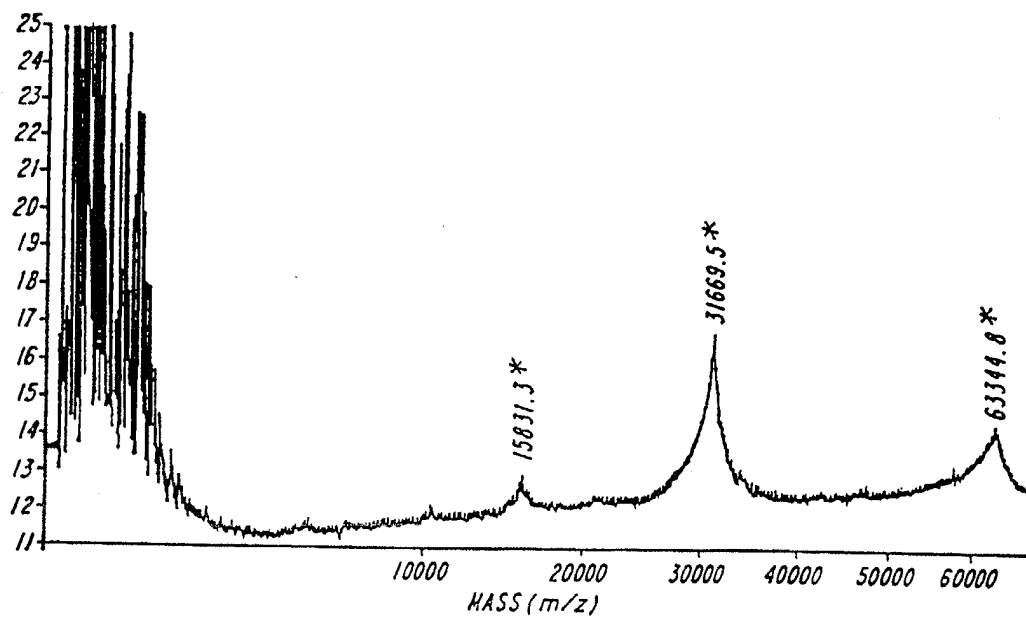


FIG. 40A

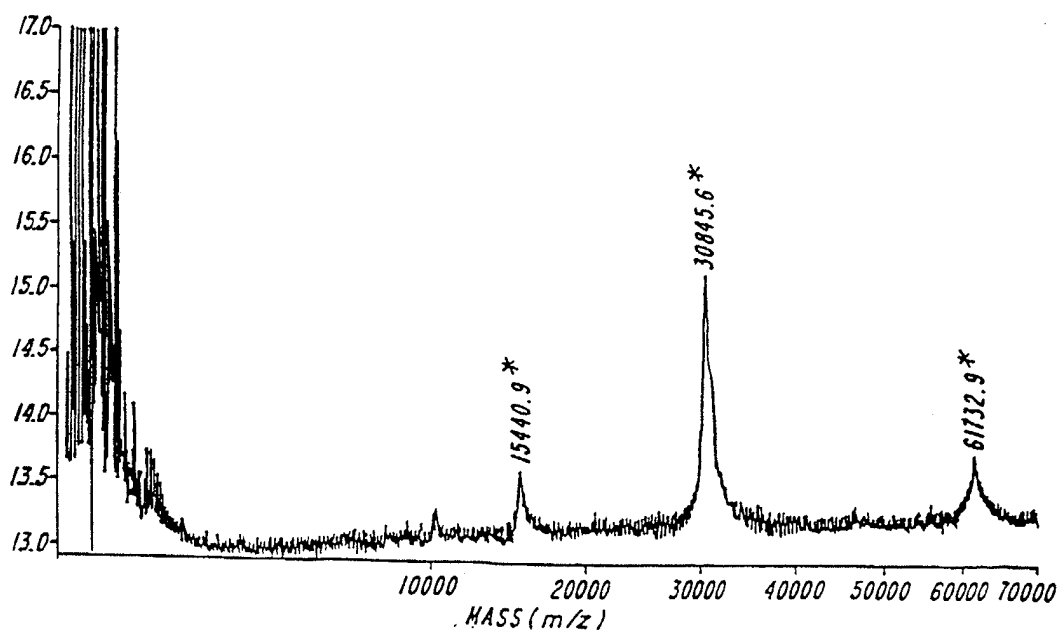


FIG. 40B

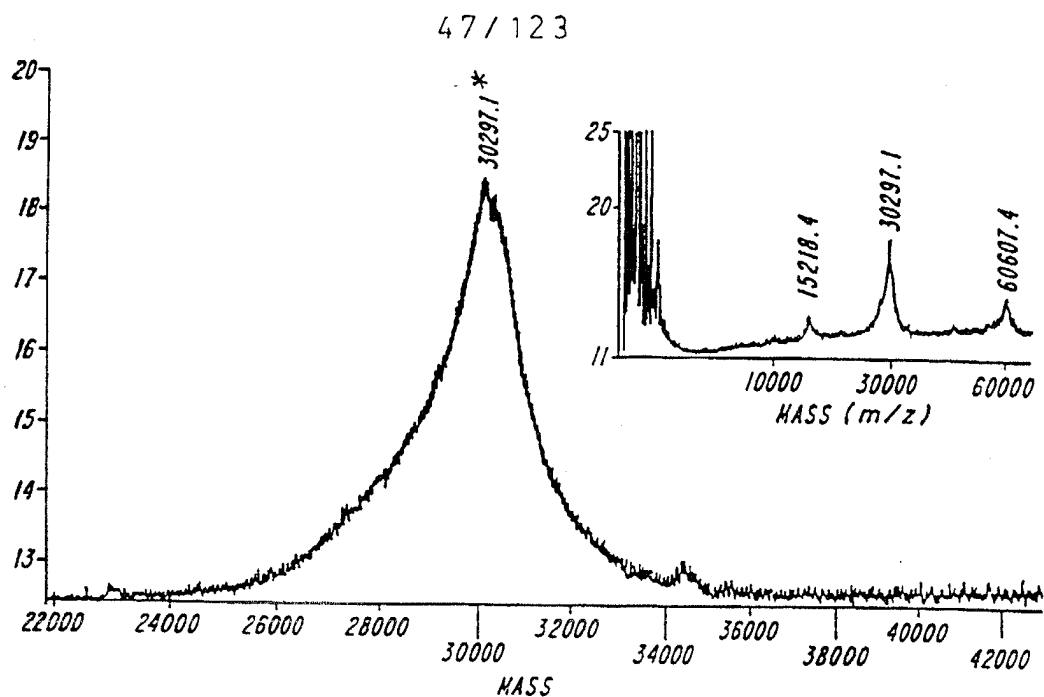


FIG. 41A

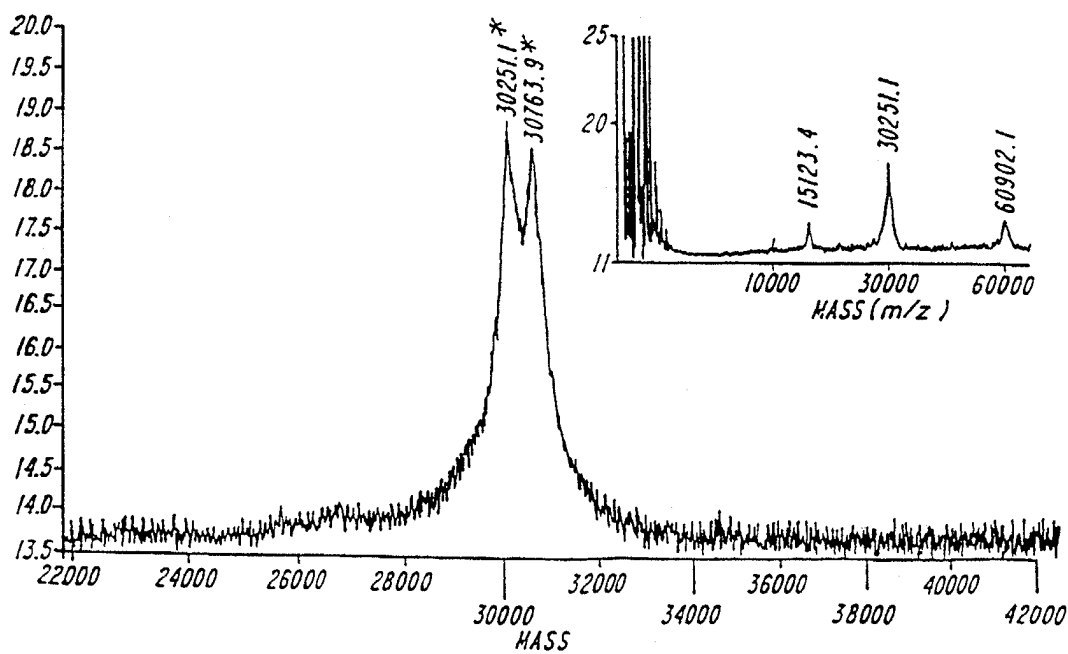


FIG. 41B

48/123

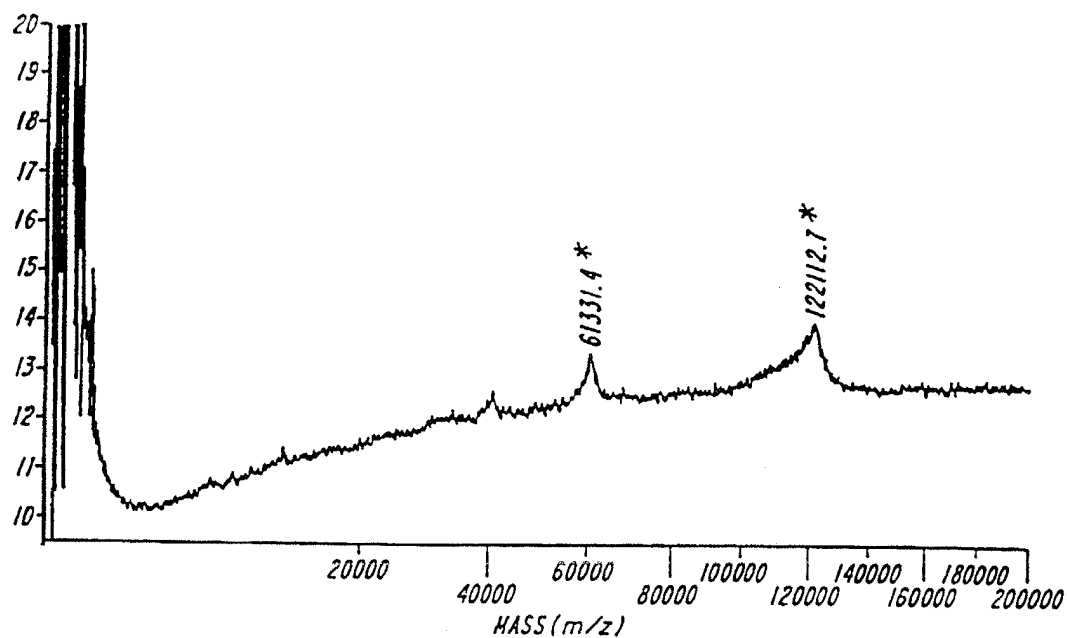


FIG. 42A

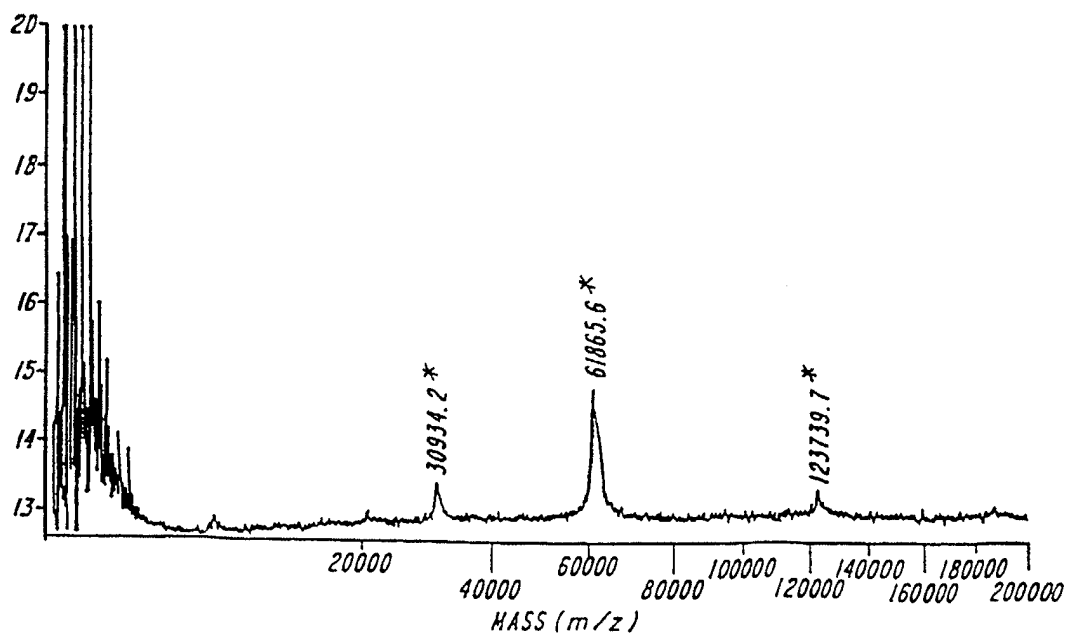


FIG. 42B

49 / 123

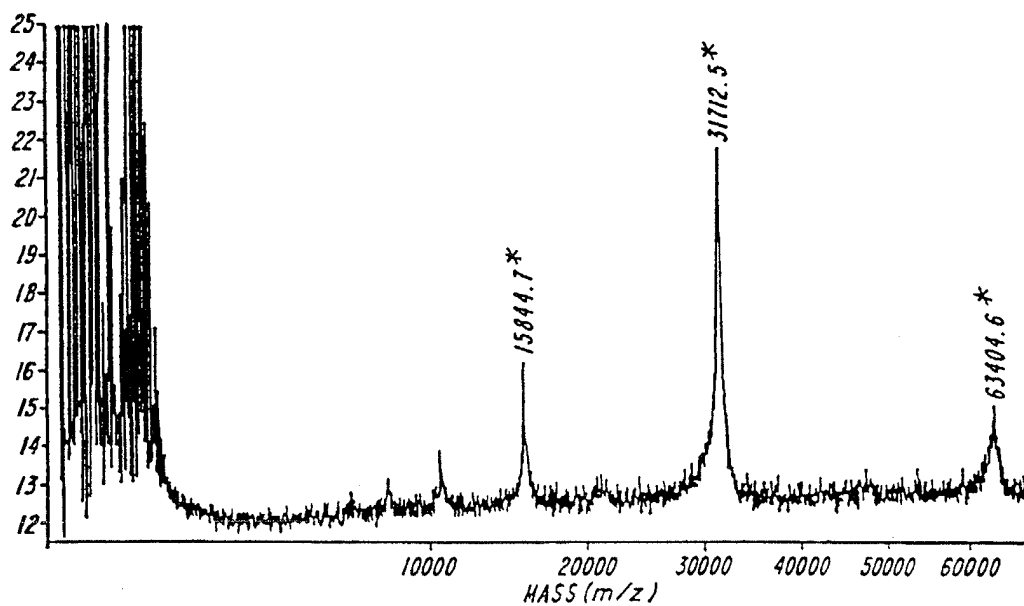


FIG. 43A

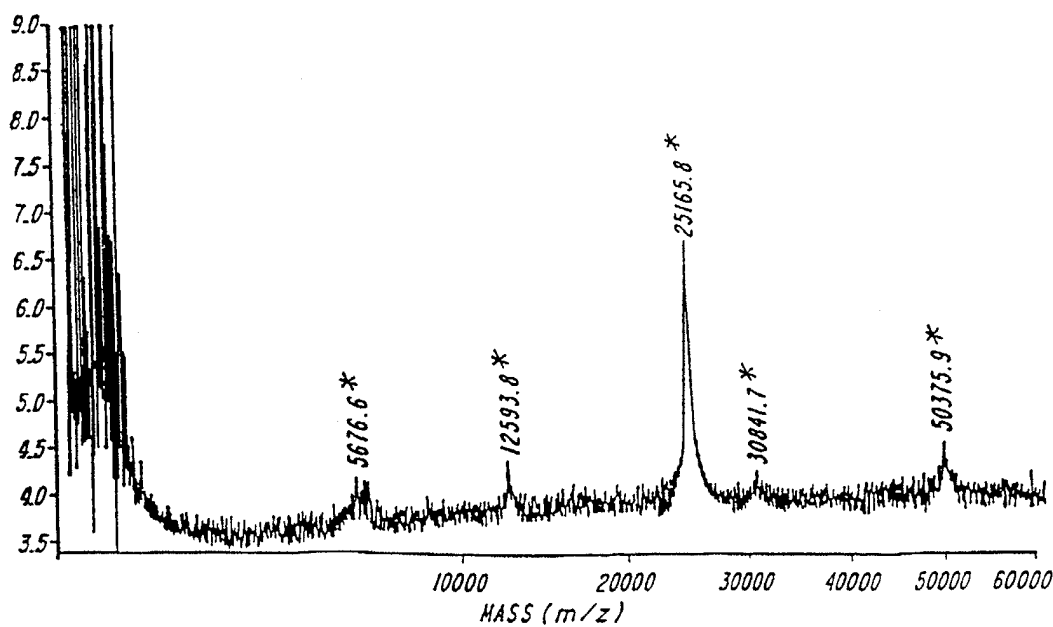


FIG. 43B

50 / 123

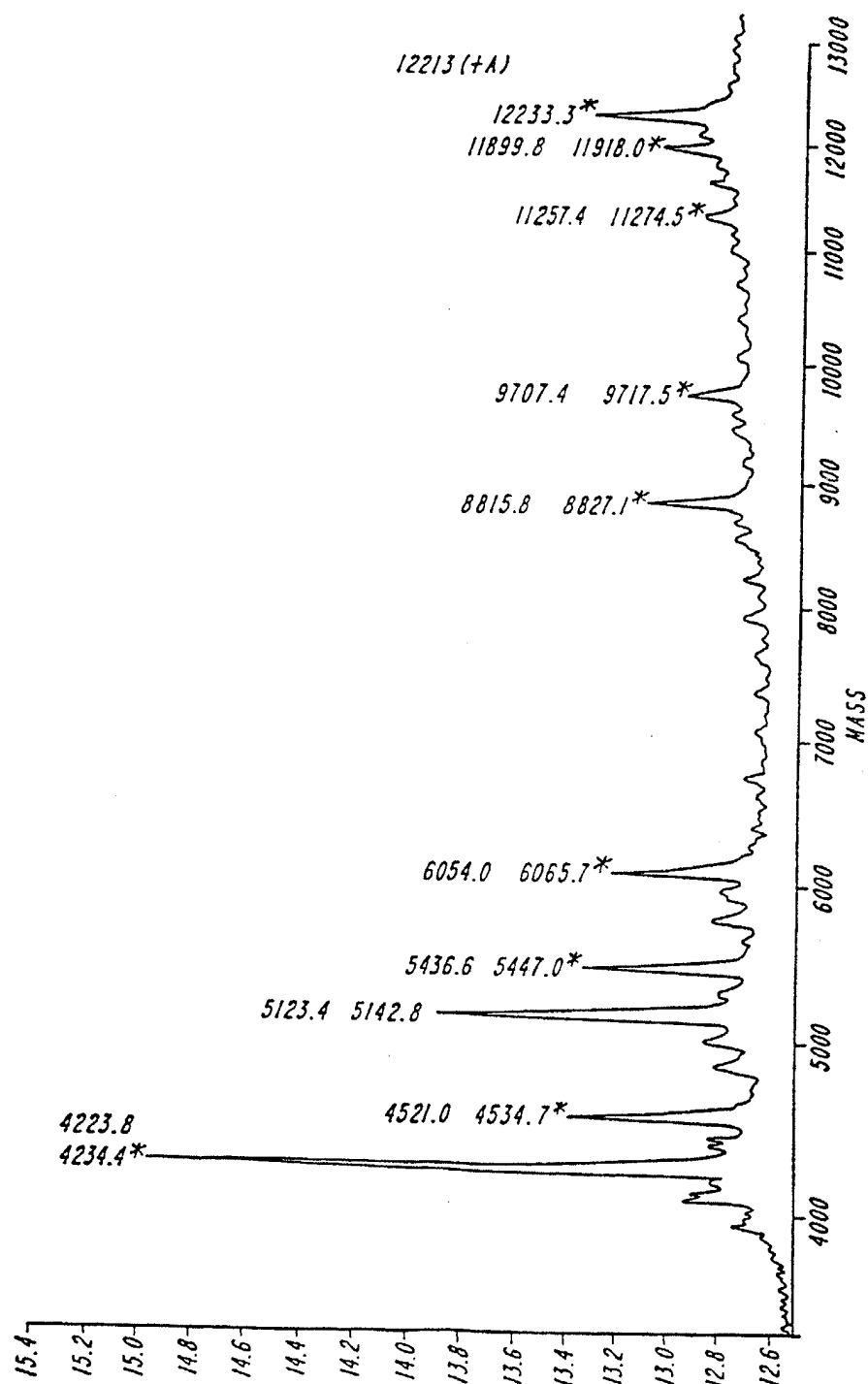


FIG. 44A

51 / 123

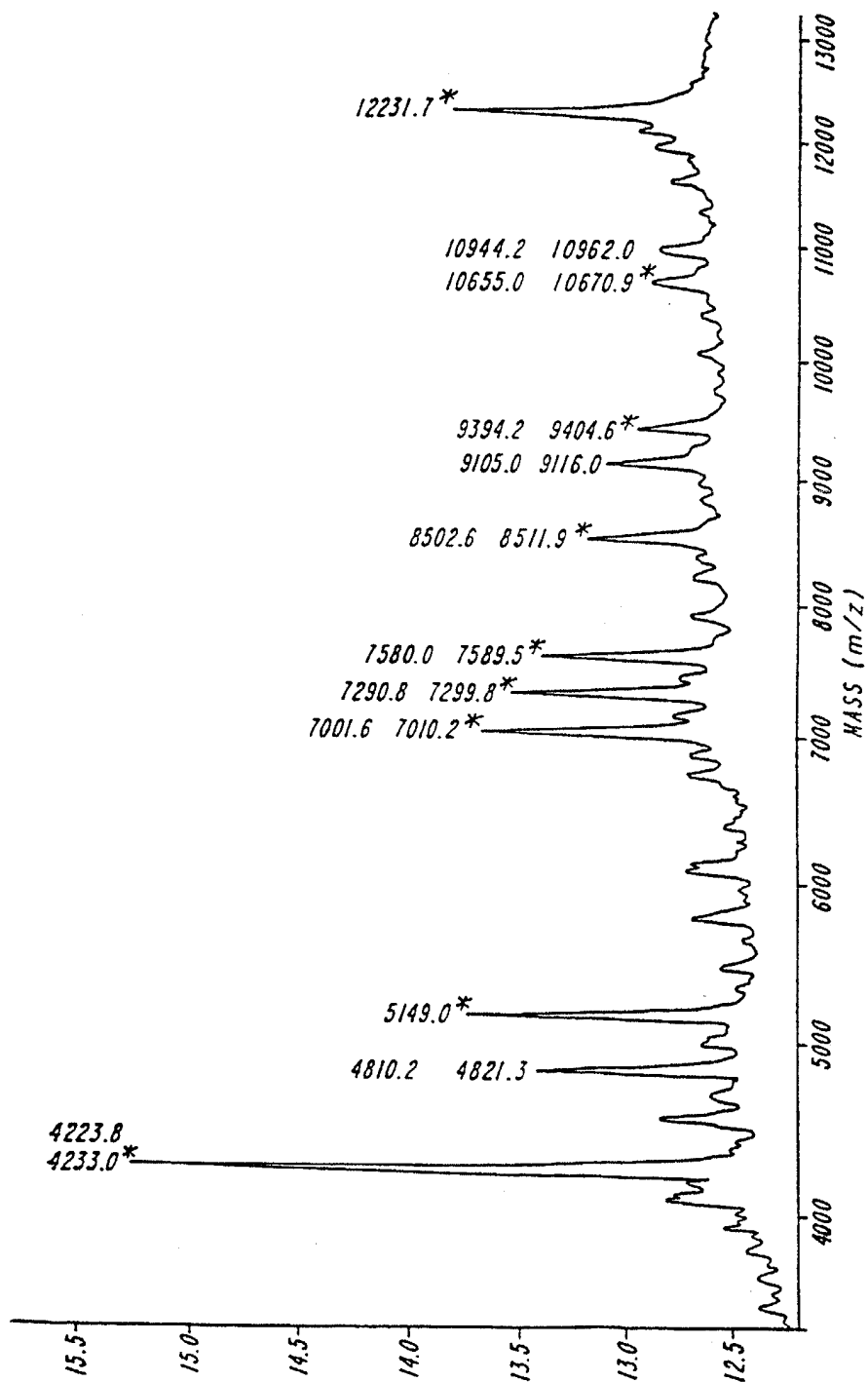


FIG. 44B

52/123

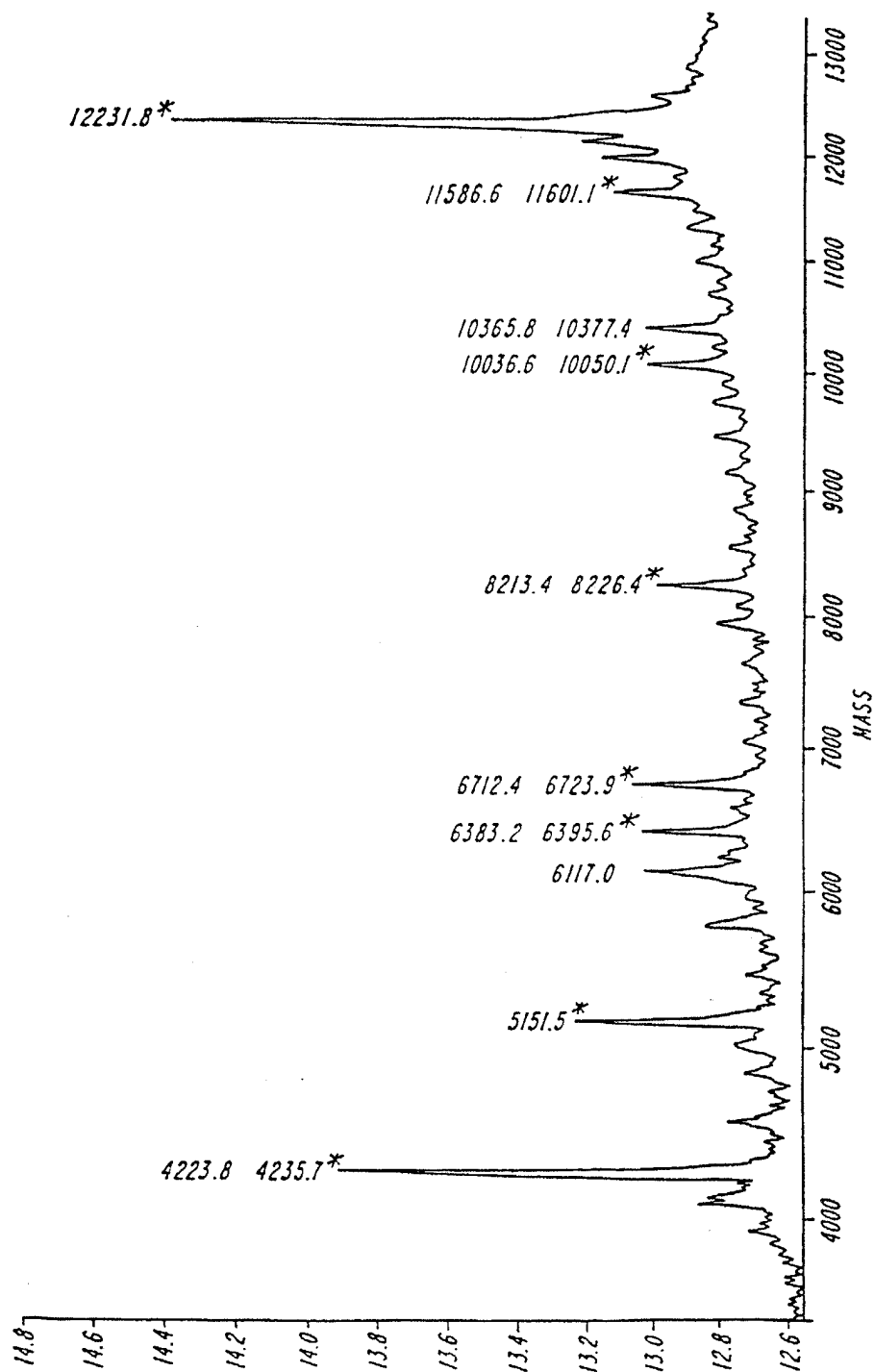


FIG. 44C

53/123

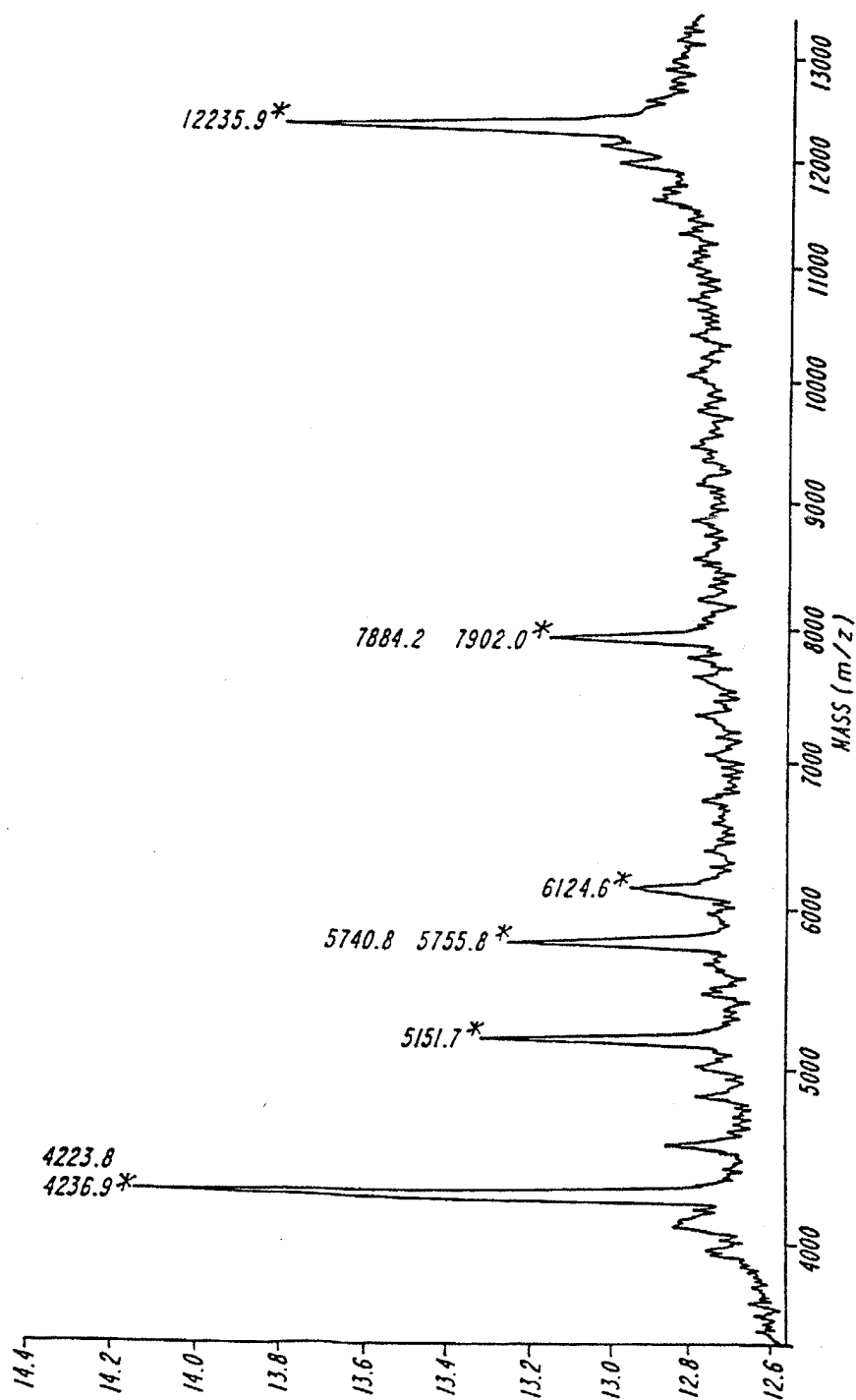


FIG. 44D

54 / 123

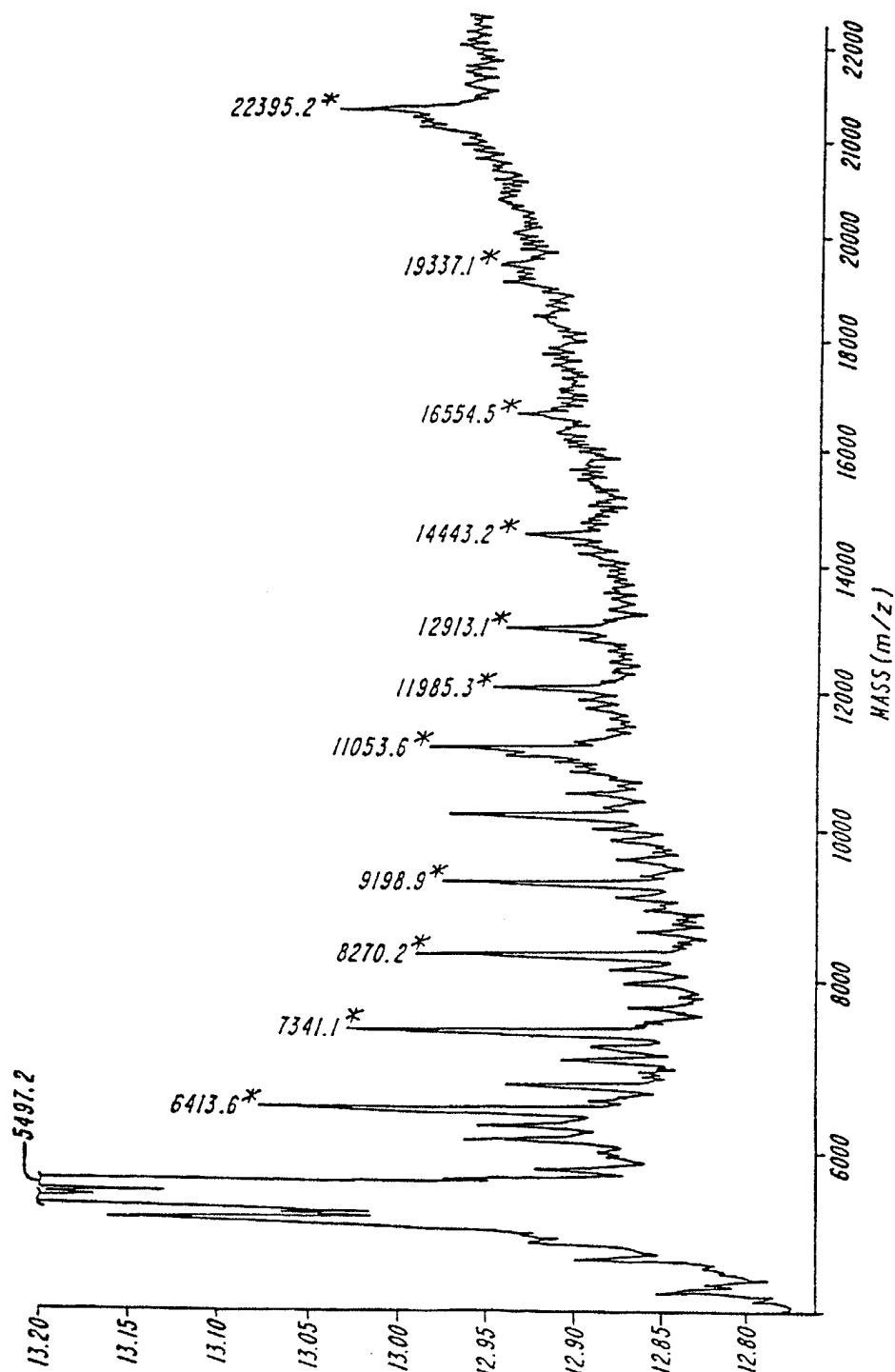


FIG. 45

55 / 123

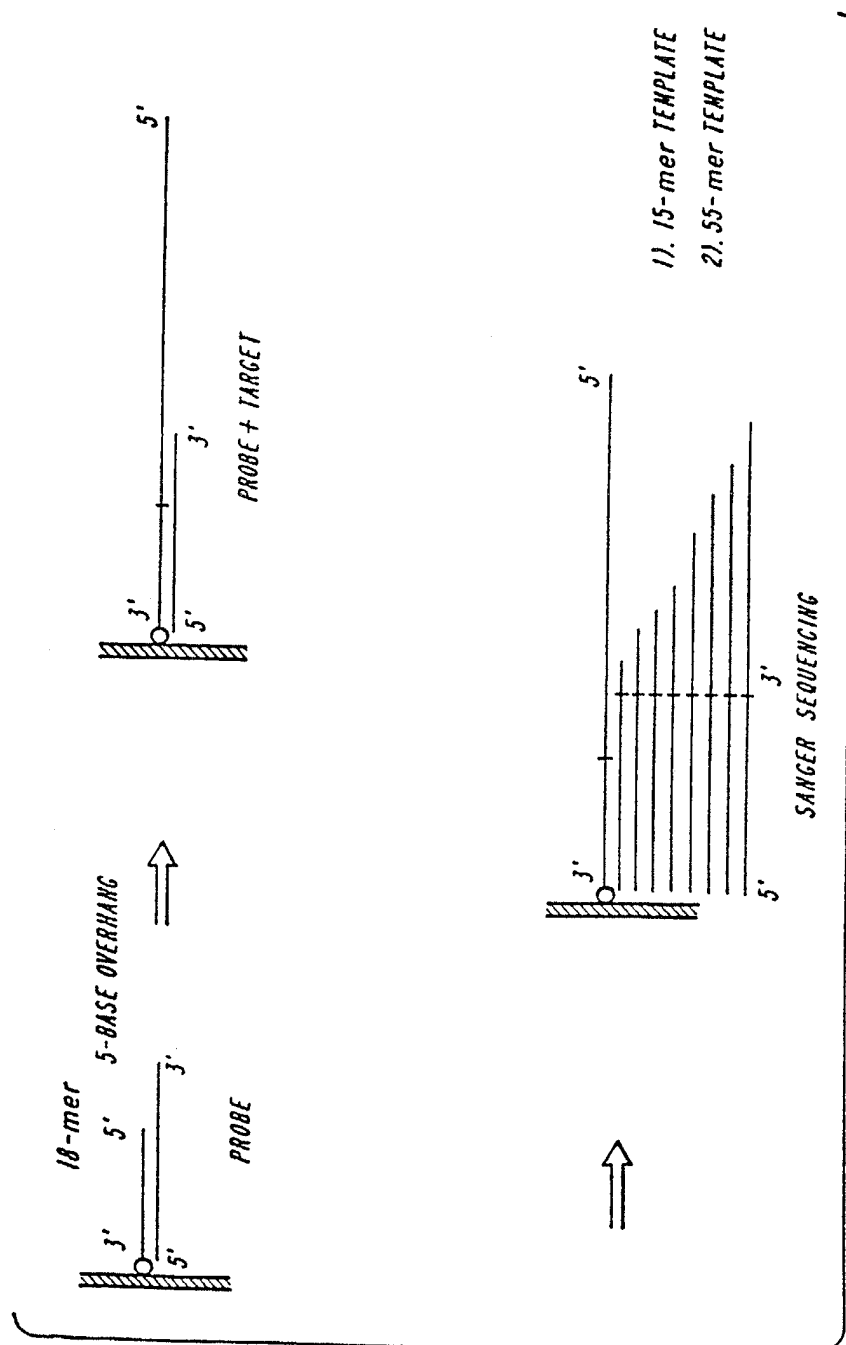


FIG. 46

56 / 123

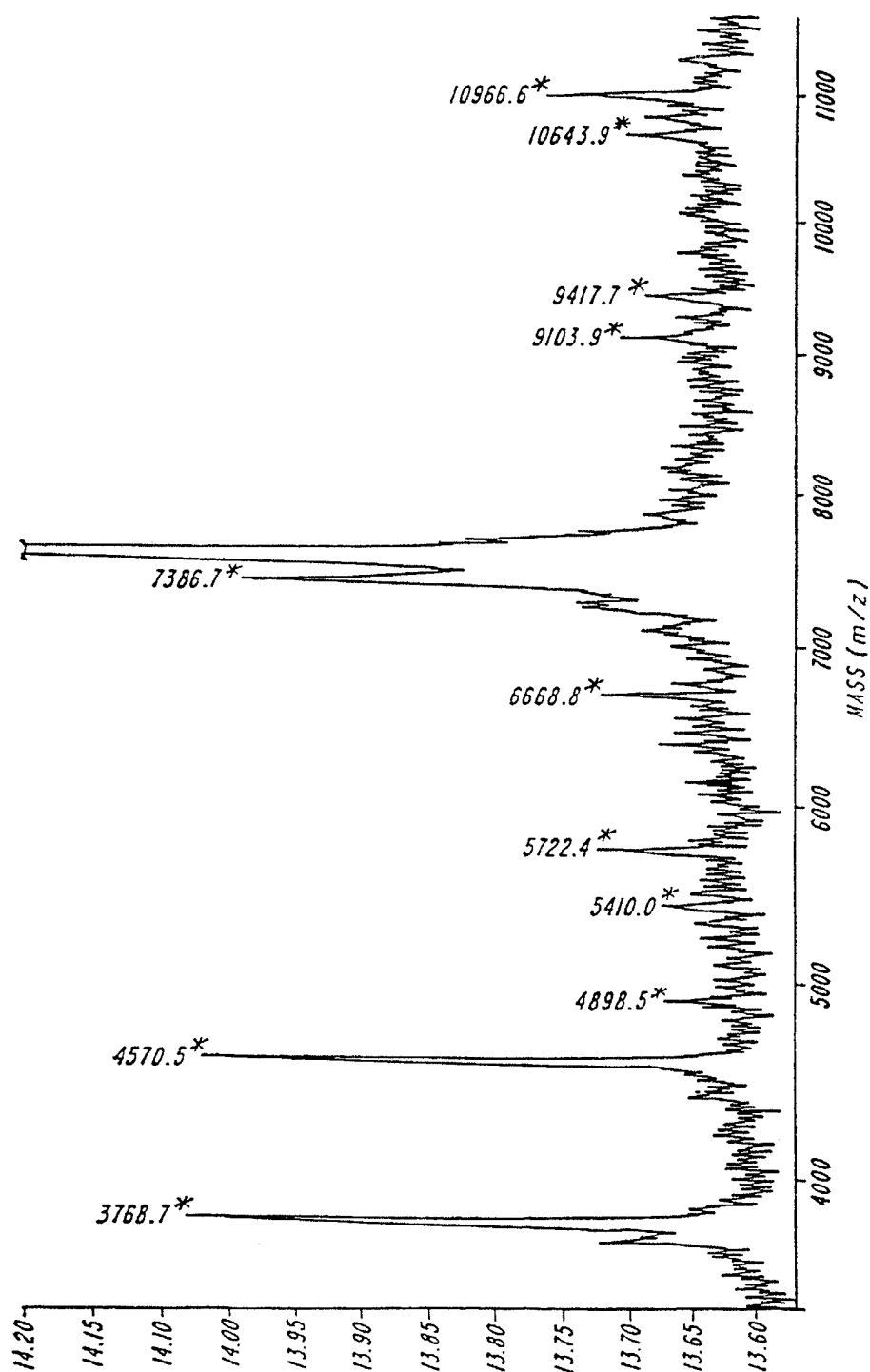


FIG. 47A

SUBSTITUTE SHEET (RULE 26)

57/123

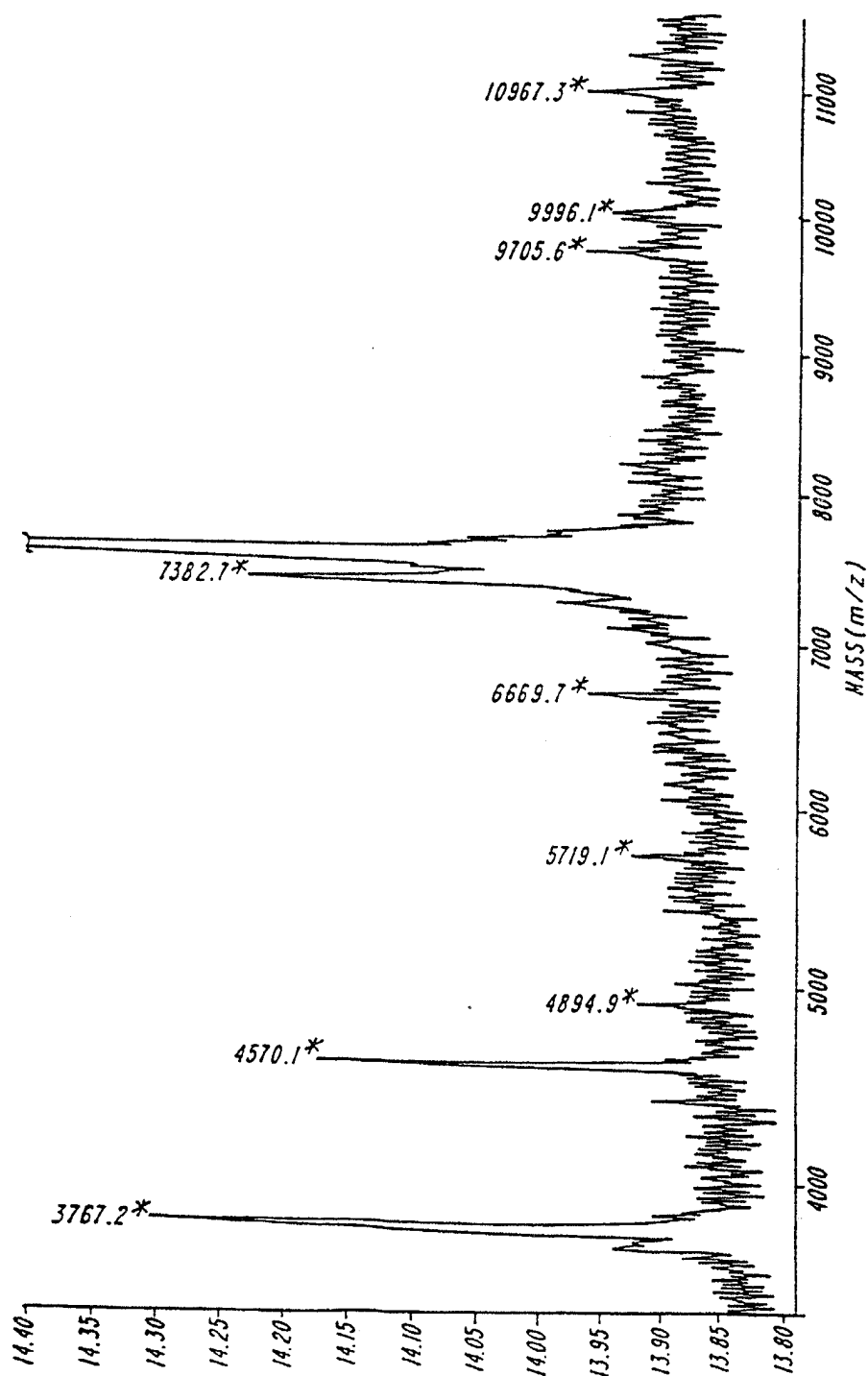


FIG. 47B

58/123

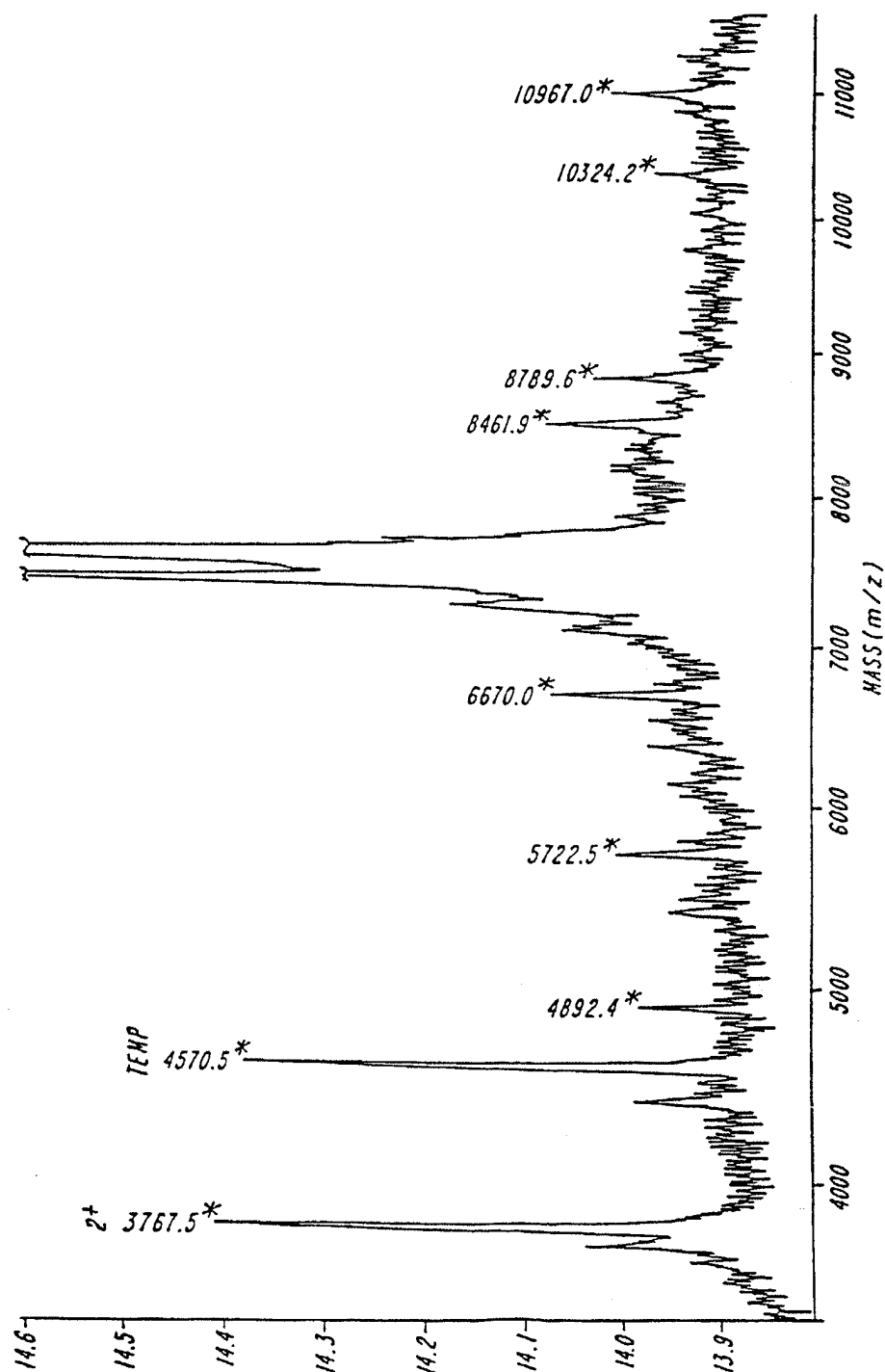


FIG. 47C

59/123

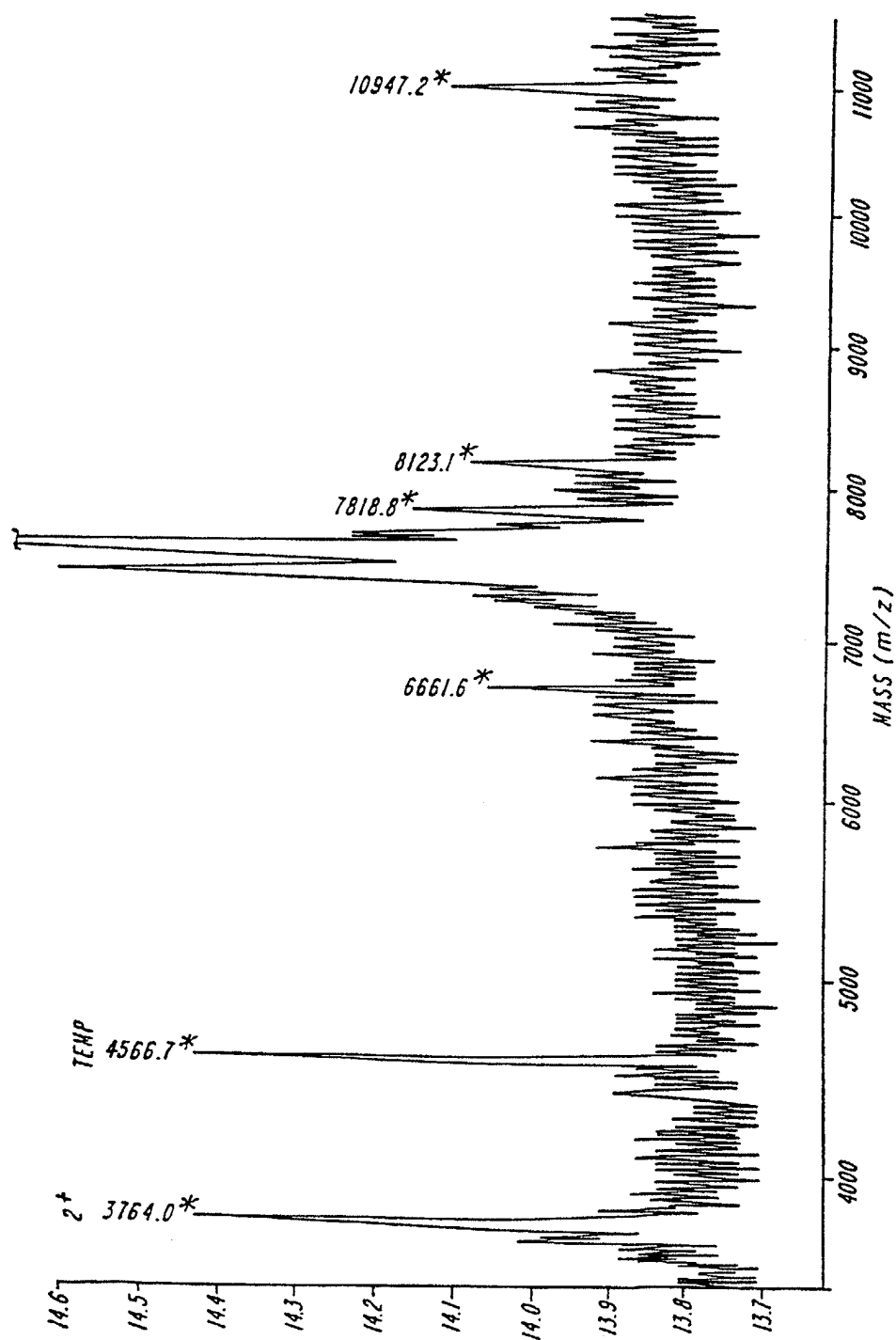
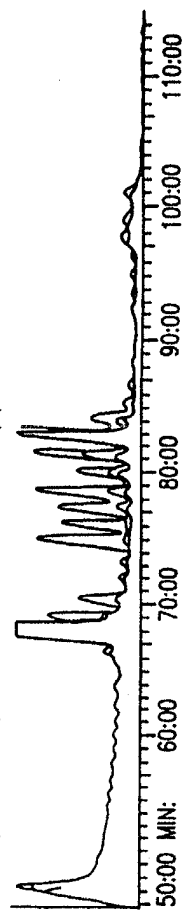
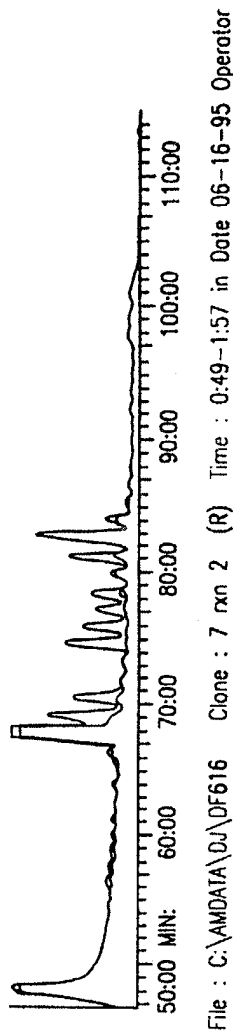
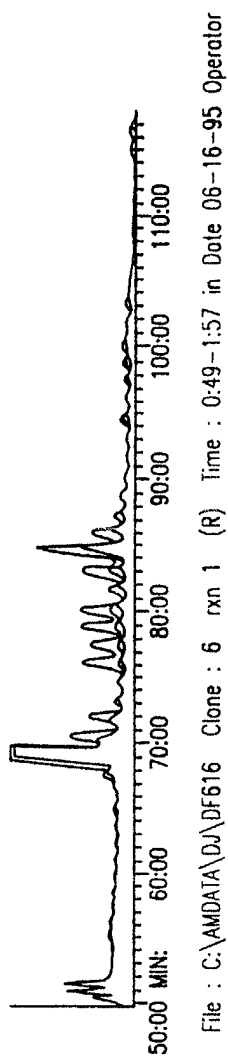
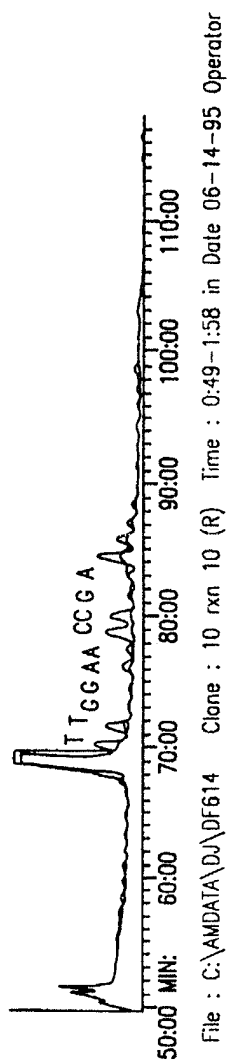


FIG. 47D

60/123



61/123

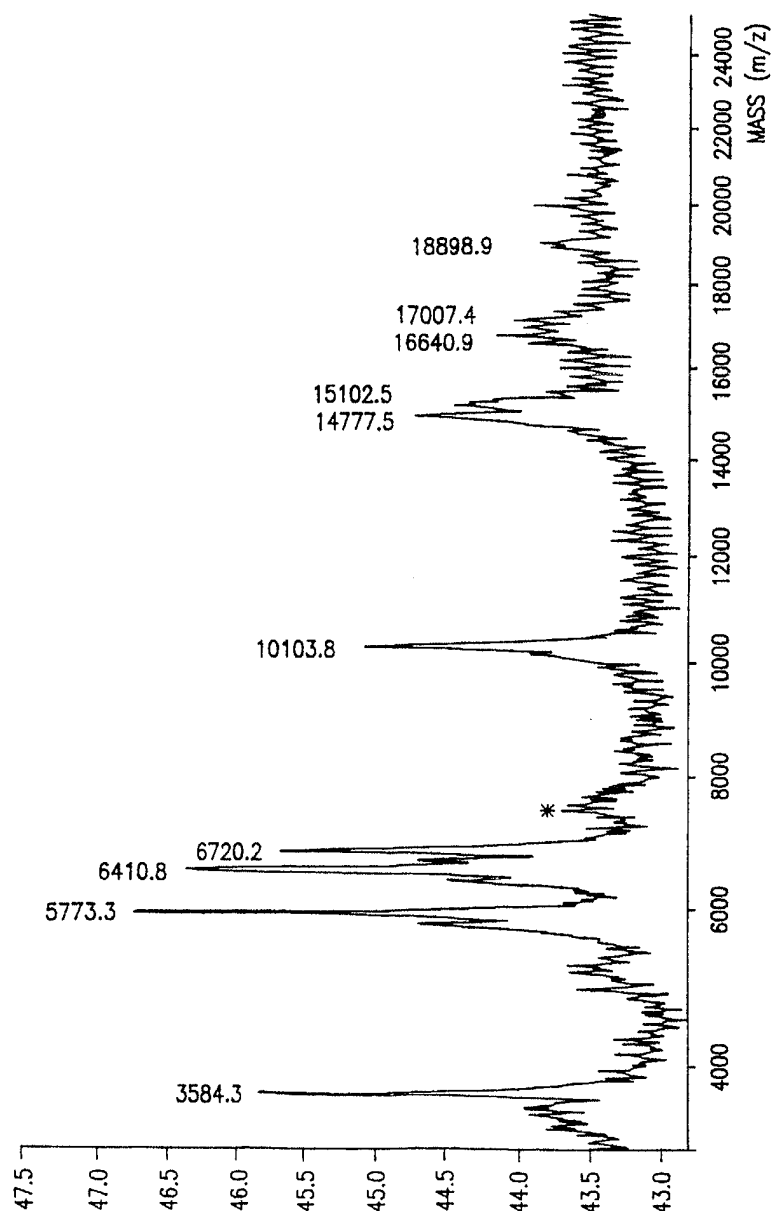


FIG. 49A

62 / 123

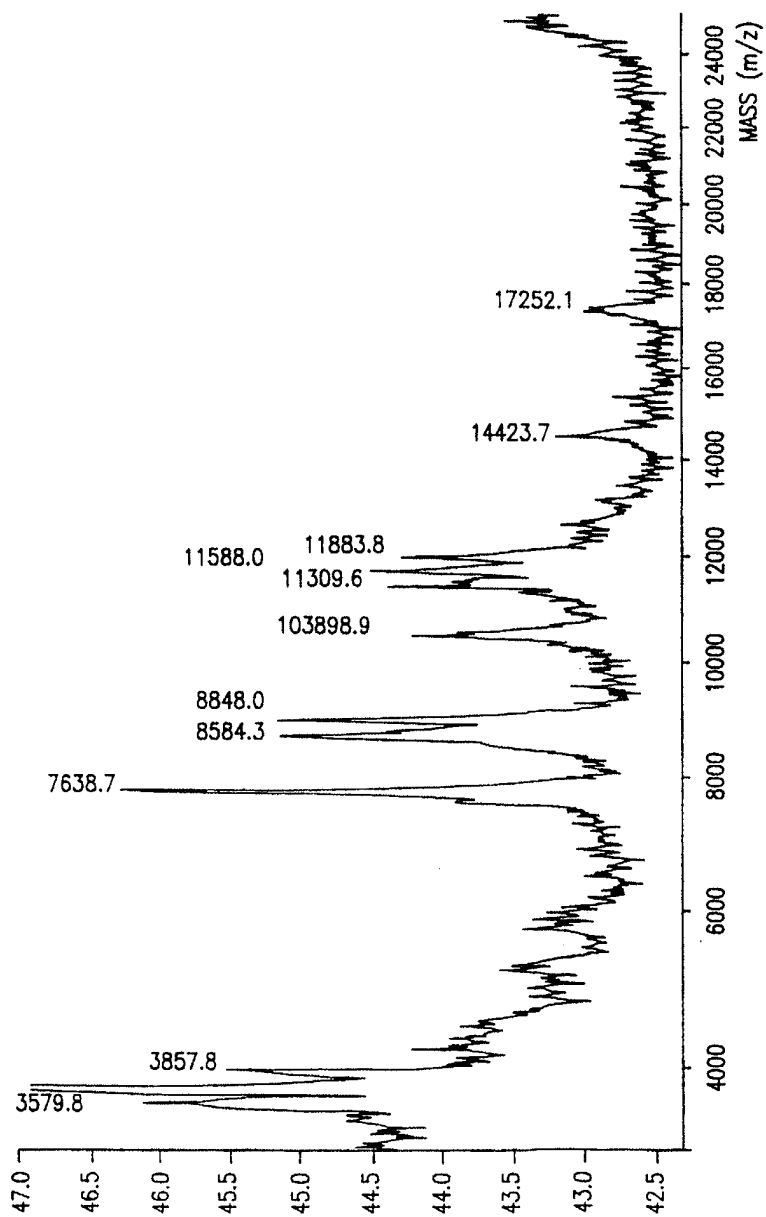


FIG. 49B

63 / 123

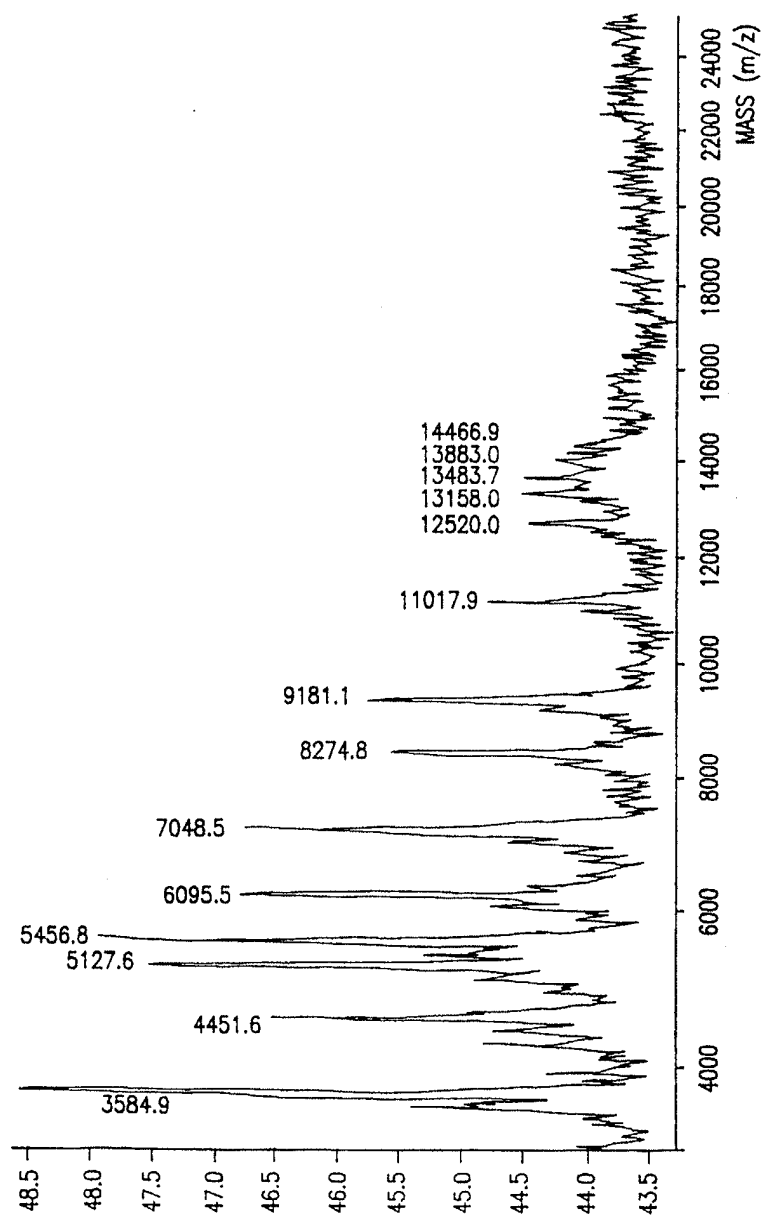


FIG. 49C

64 / 123

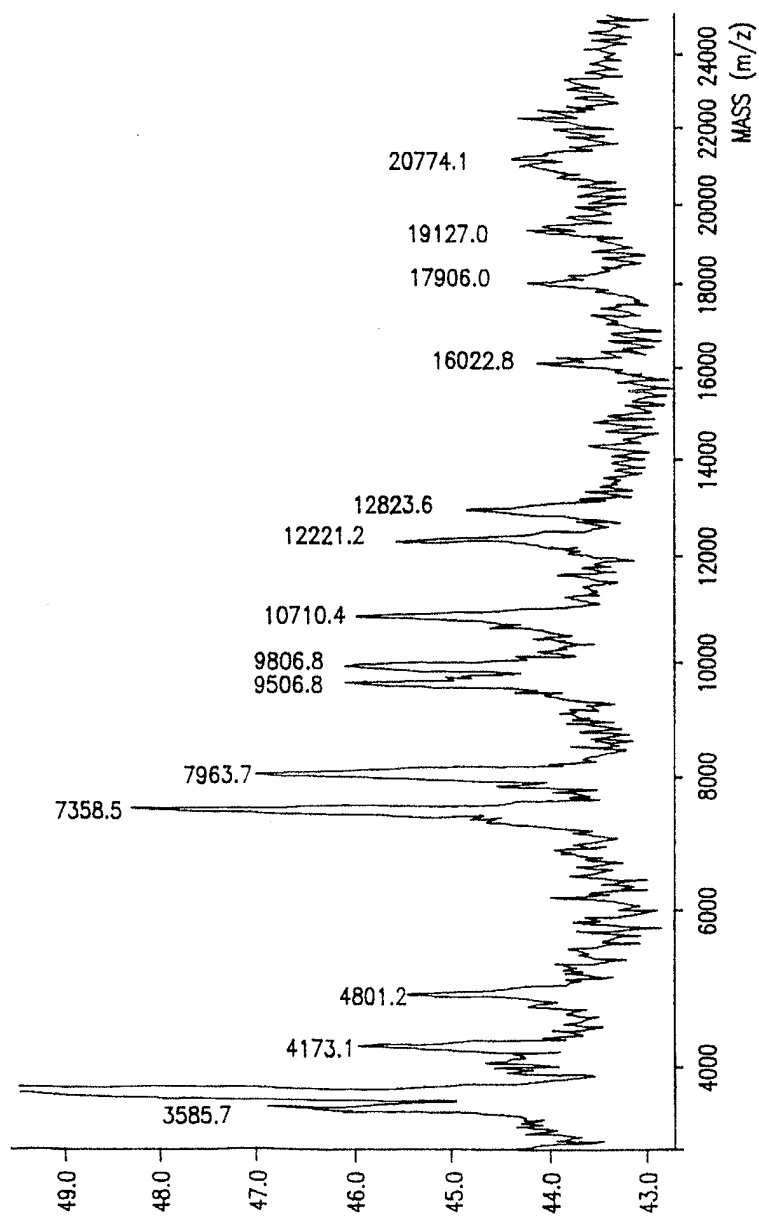


FIG. 49D

65/123

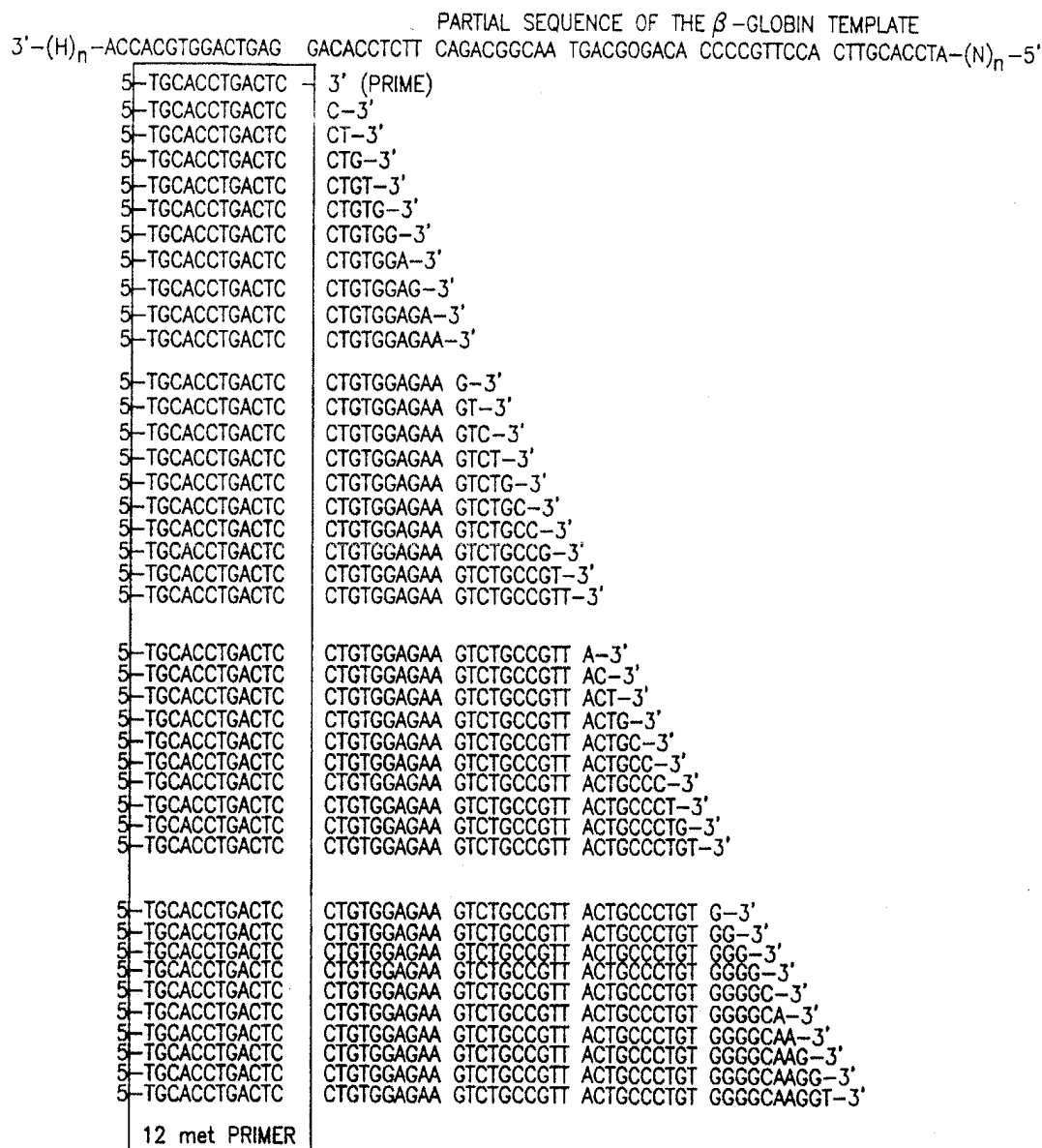


FIG. 50A

66/123

REACTION STOPPED WITH			
ddATP	ddCTP	ddGTP	ddTTP
3581.4 da	3581.4 da 3854.6 da	3581.4 da	3581.4 da
		4488.0 da	4158.9 da
		5120.4 da 5448.6 da	4791.2 da
5760.8 da		6089.0 da	
6401.2 da 6713.4 da		7041.6 da	7344.8 da
	7634.0 da	8267.4 da	7938.2 da
	8555.6 da 8844.8 da	9174.0 da	9477.2 da 9781.4 da
10094.6 da	10382.8 da	11016.2 da	10687.0 da
	11304.4 da 11593.6 da 11652.8 da	12516.2 da	12187.0 da 12819.4 da
		13148.6 da 13476.8 da 13805.0 da 14133.2 da	
14734.6 da 15146.8 da	14421.4 da	15375.0 da 15703.2 da	16006.4 da

FIG. 50B

67 / 123

SEQUENCE OF THE AMPLIFIED 209 bp PCR-PRODUCT OF THE β -GLOBIN GENE

FORWARD PRIMER: β 2
CATTGCTTC TGACACAACT GTGTTCACTA GCAACCTCAA ACAGACACCA
12mer PRIMER
TGGTGCACCT GACTGCTGTG GAGAAGTCTG CCGTTACTGC CCTGTGGGGC
AAGGTGAACG TGGATGAACT TGGTGGTGAG GCCCTGGGCA GGTGGTATC
AAGGTTACAA GACAGGTTTA AGGAGACCAA TAGAACTGG GCATGTGGAG
ACAGAGAAG
REVERSE PRIMER β 11

FIG. 51

68/123

ALLELE	NUMBER OF REPEATS TRUNCATED	THEORETICALLY CALCULATED		MOLECULAR MASS ddG AND ddC
		ddG	ddC	
8xAAAT		19440.60	11643.60	11643.60
9xAAAT		15718.20	21033.60	15718.20
10xAAAT		16959.00	22774.40	16959.00
11xAAAT		18199.80	23515.20	18199.80
12xAAAT		19440.60	24756.00	19440.60
13xAAAT		20681.40	25996.80	20681.40
13xAAAT		21922.20	27237.60	21922.20

FIG. 52

69/123

FIG. 53A

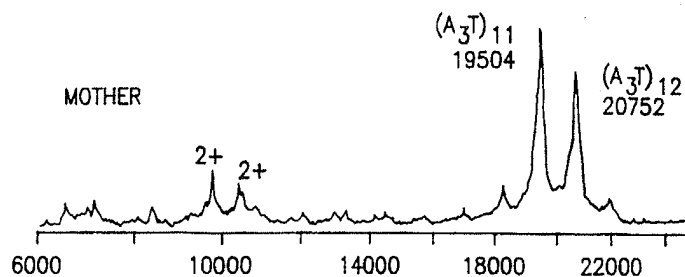


FIG. 53B

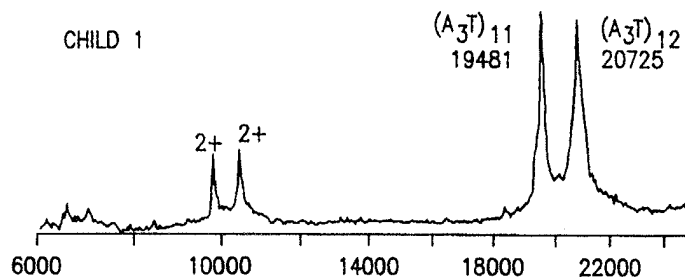


FIG. 53C

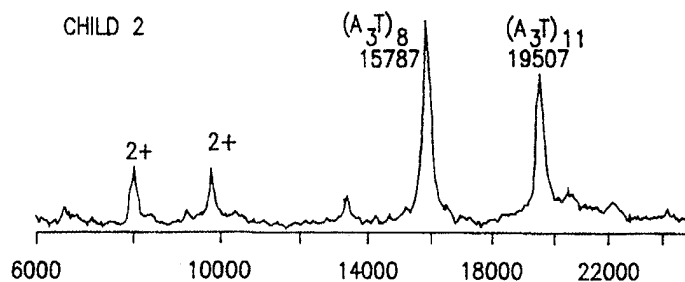


FIG. 53D

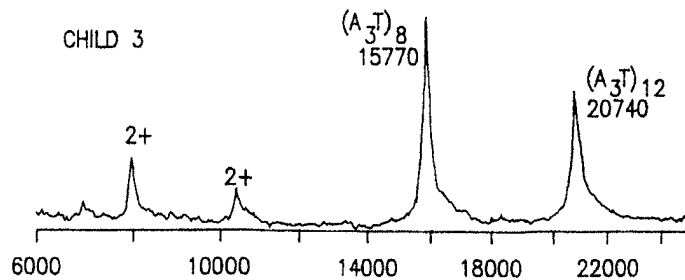
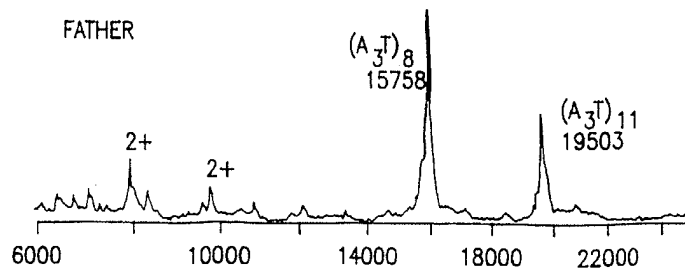
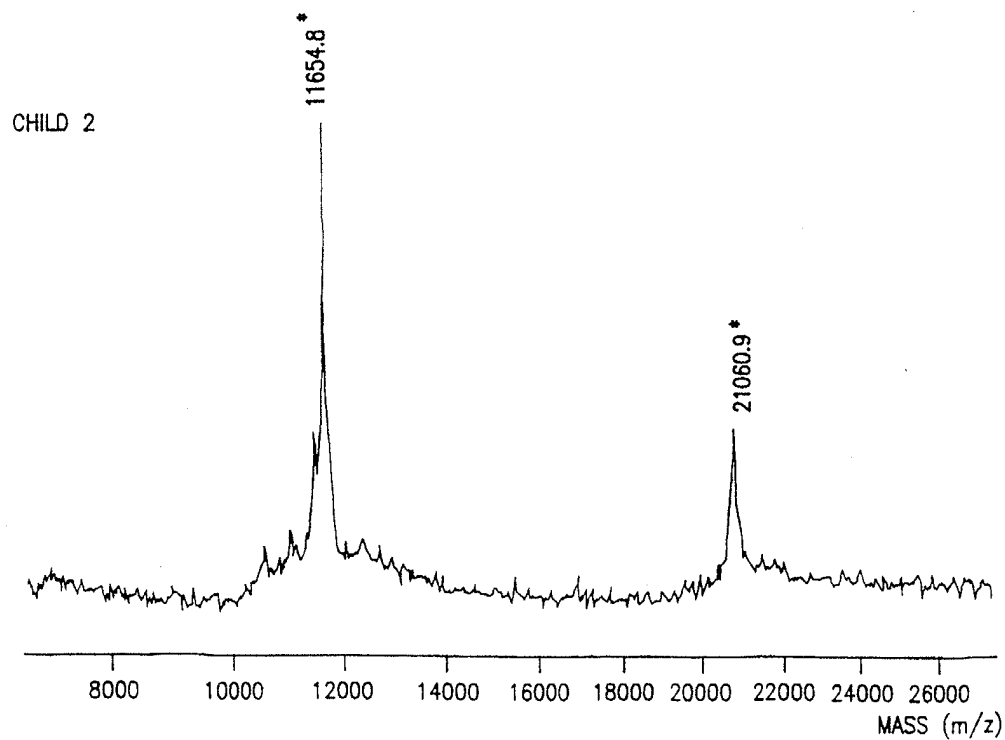
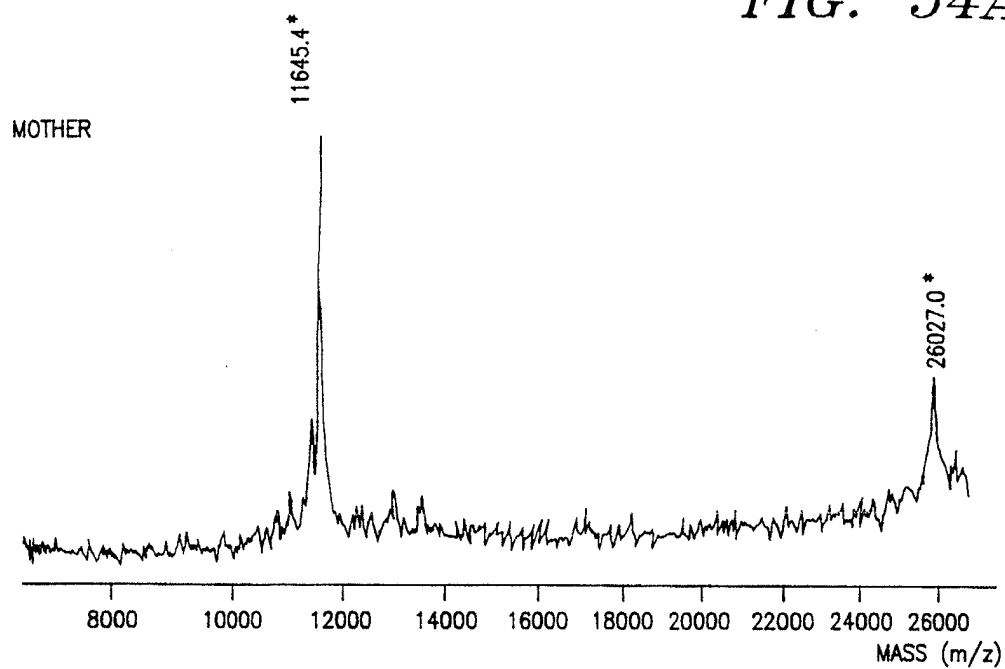


FIG. 53E



SUBSTITUTE SHEET (RULE 26)

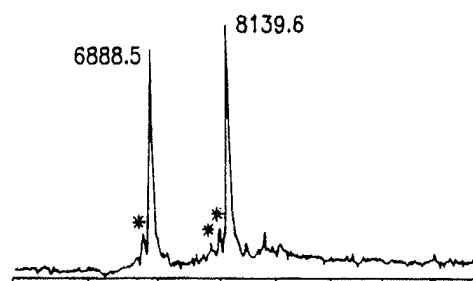
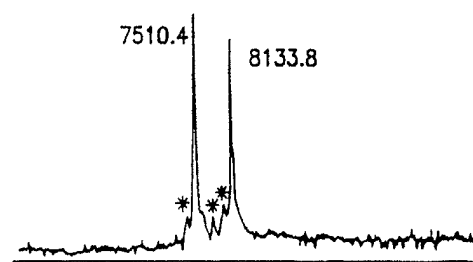
70/123

FIG. 54A

SUBSTITUTE SHEET (RULE 26)

5' -GTGTGTGTGTGTGTGTTTT (TT) (TT) AACAGGGATTGGGAATTATTGAGA-3'
PRIMER TTGCCCTAAACCCCTT (4448.0)
T5 ALLELE CAAAAA --- -- TTGCCCTAAACCCCTT (6890.6)
T7 ALLELE CAAAAA AA -- TTGCCCTAAACCCCTT (7515.0)
T9 ALLELE CAAAAA AA AA TTGCCCTAAACCCCTT (8139.4)

FIG. 55

FIG. 56A*FIG. 56B*

73/123

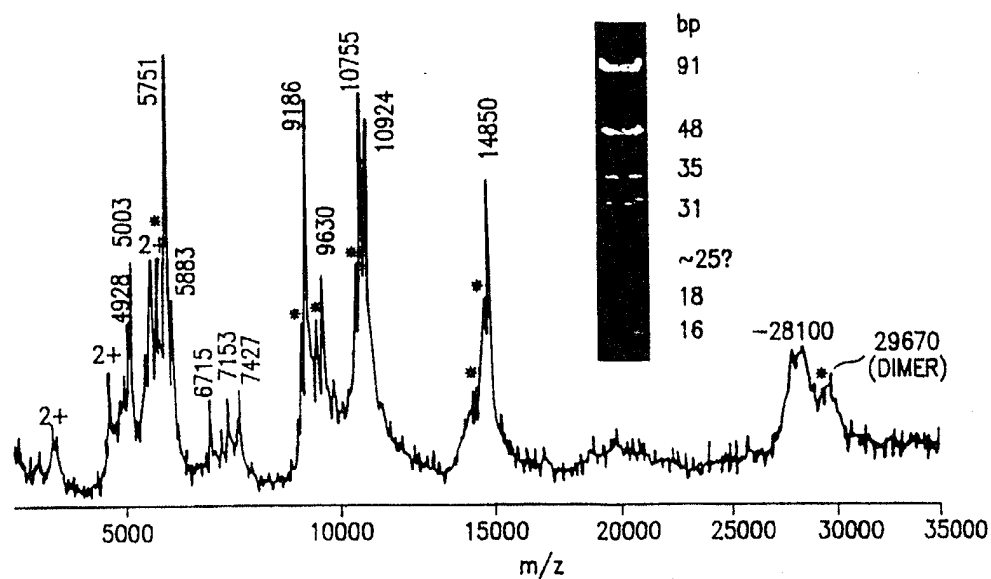


FIG. 57A

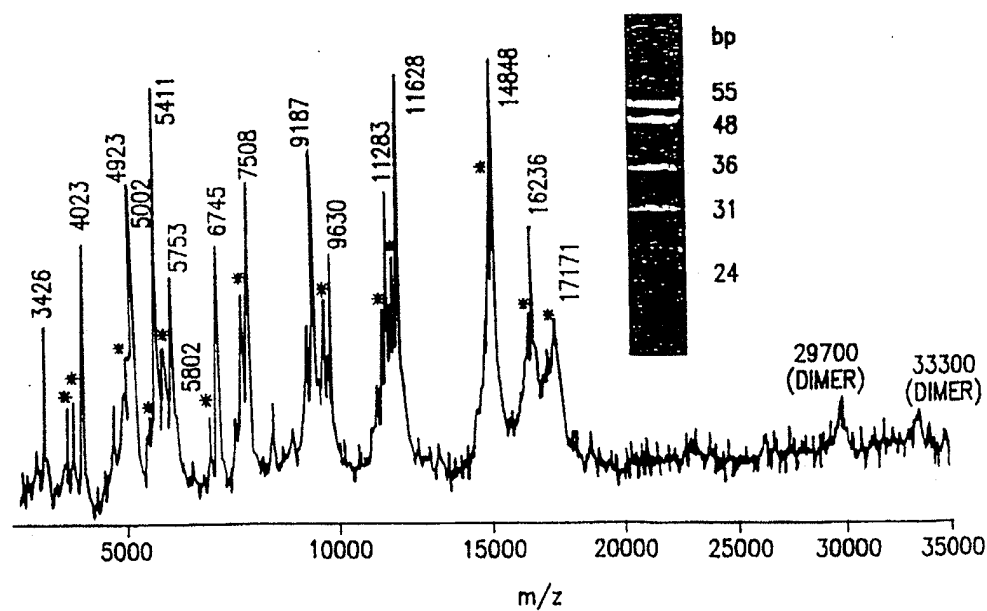


FIG. 57B

74 / 123

FIG. 58A

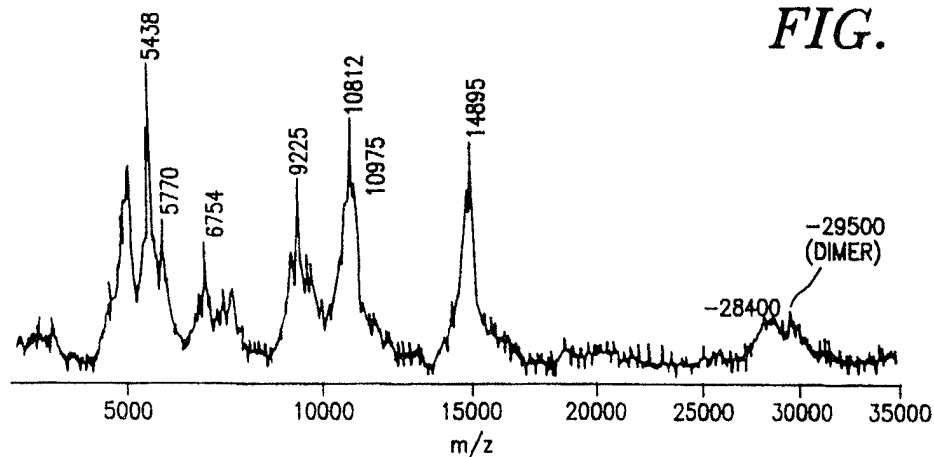


FIG. 58B

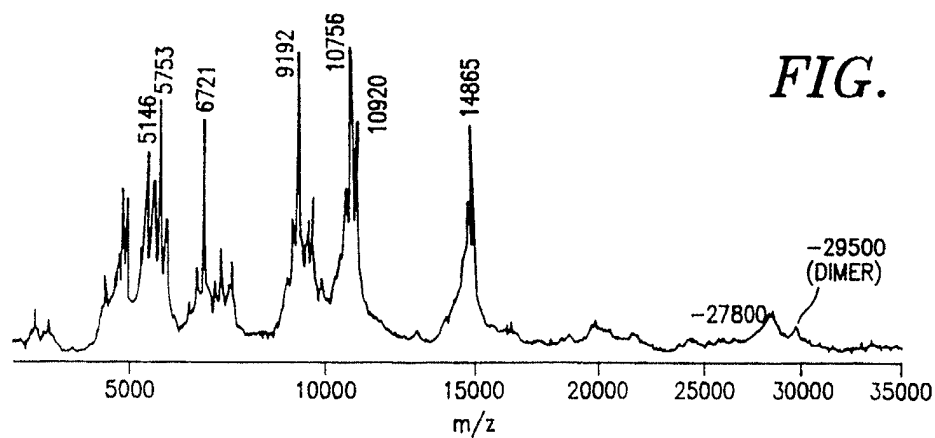
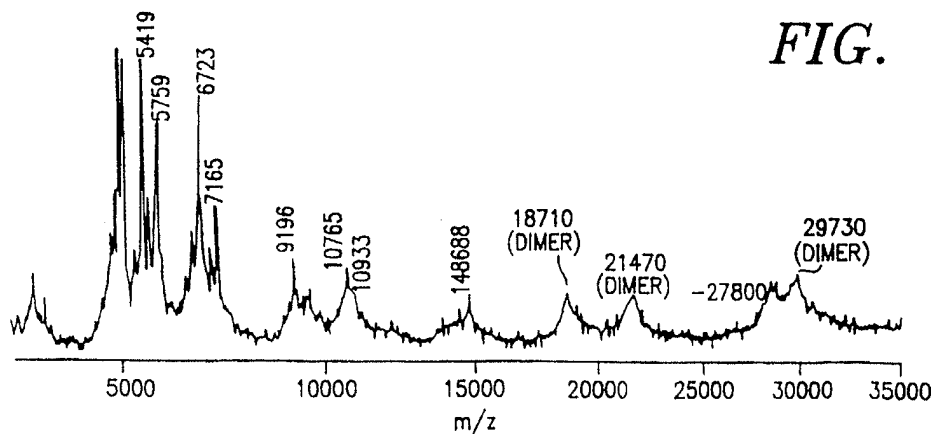


FIG. 58C



75/123

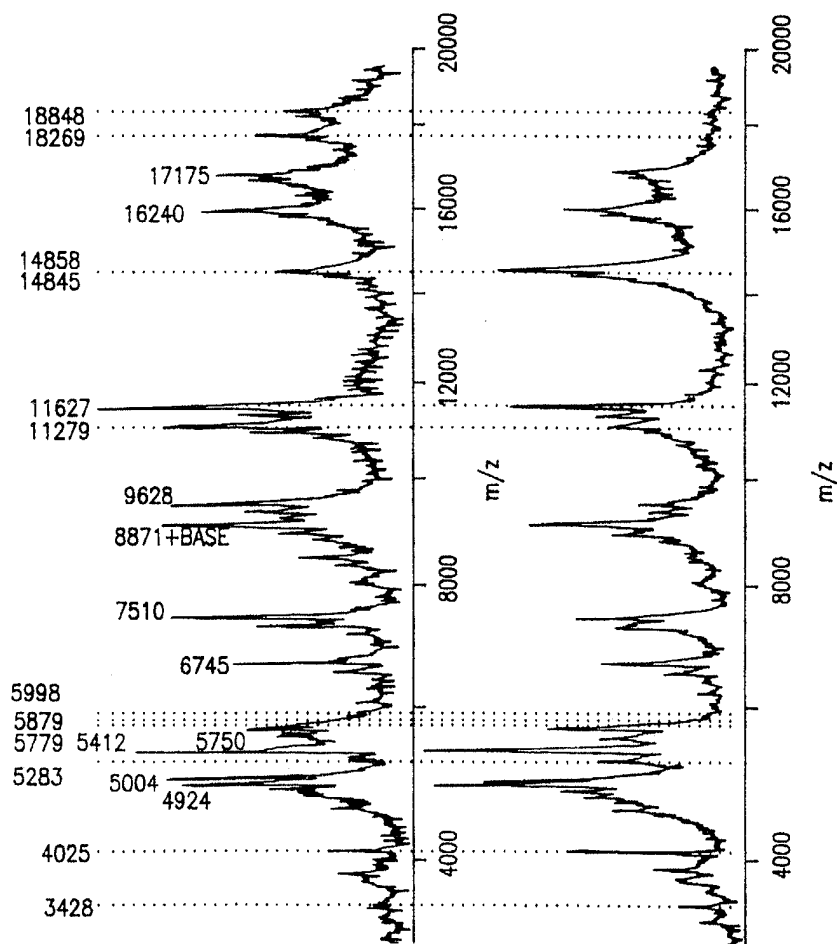


FIG. 59A

FIG. 59B

76/123

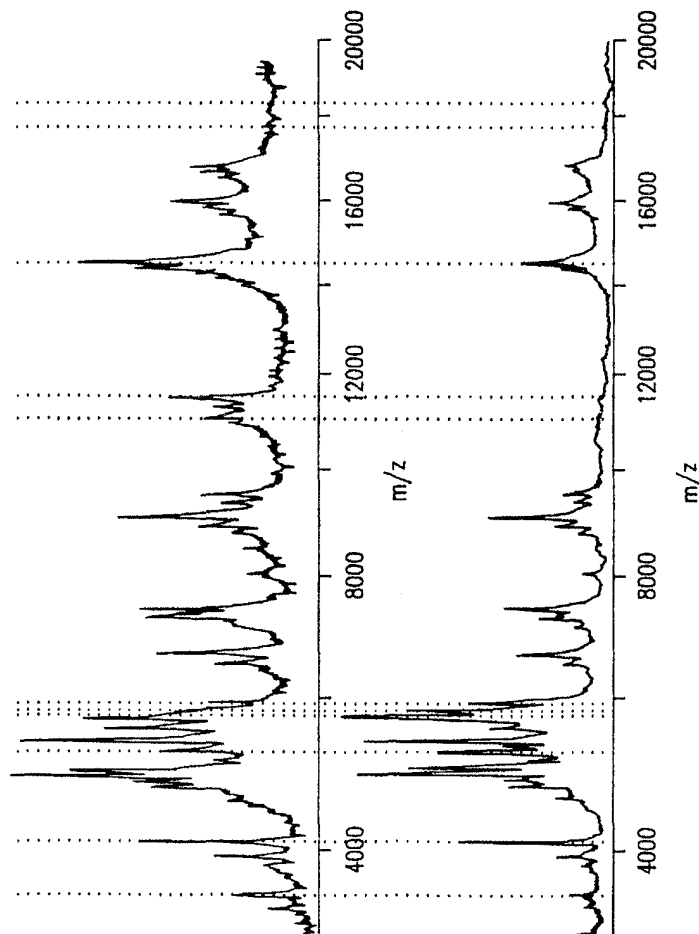


FIG. 59C

FIG. 59D

77/123

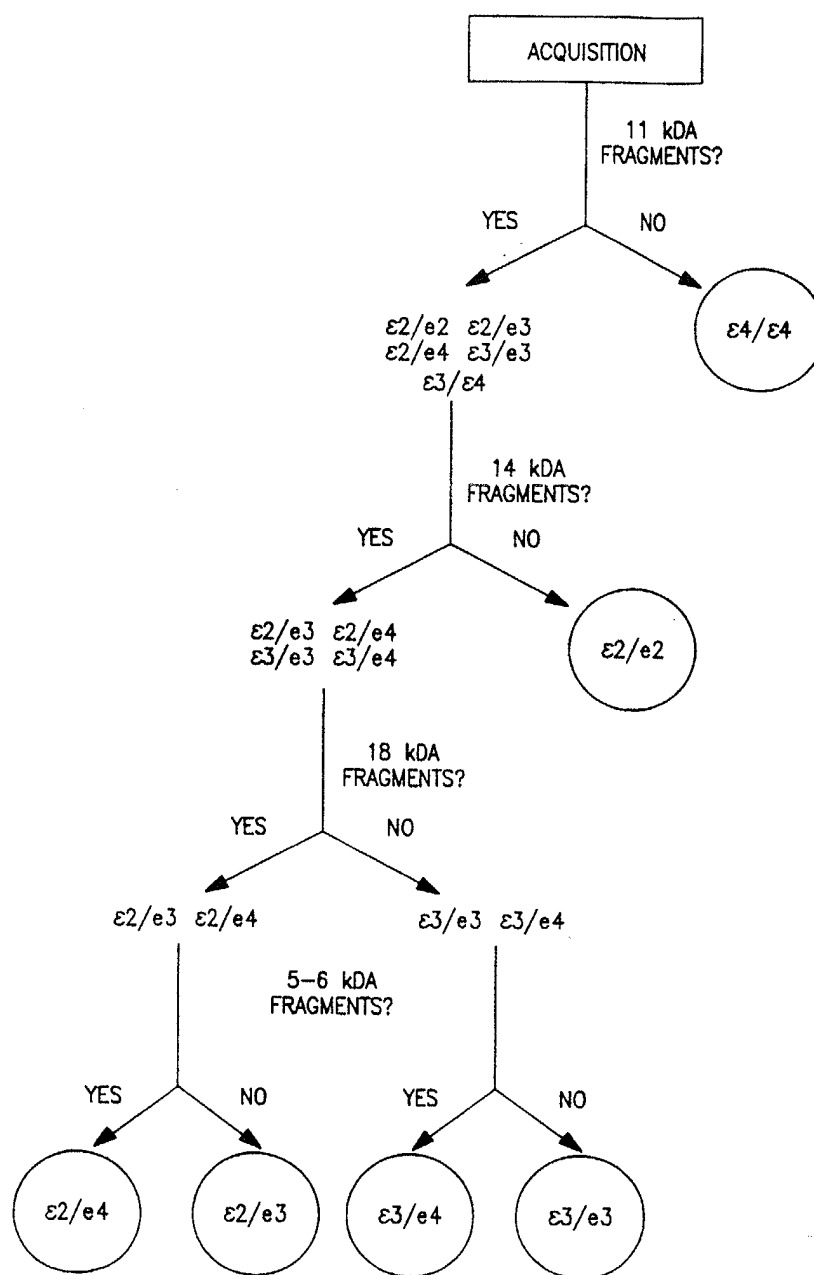
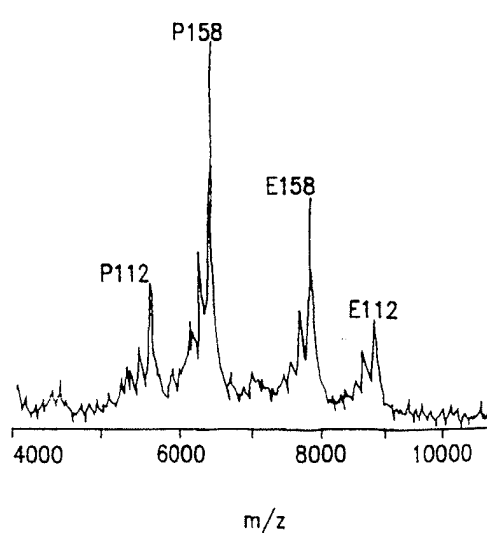
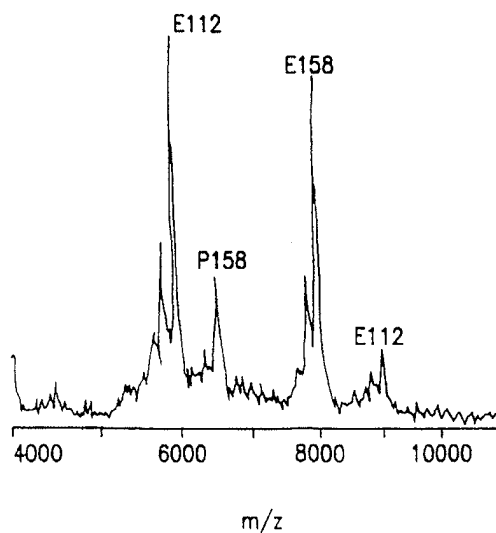
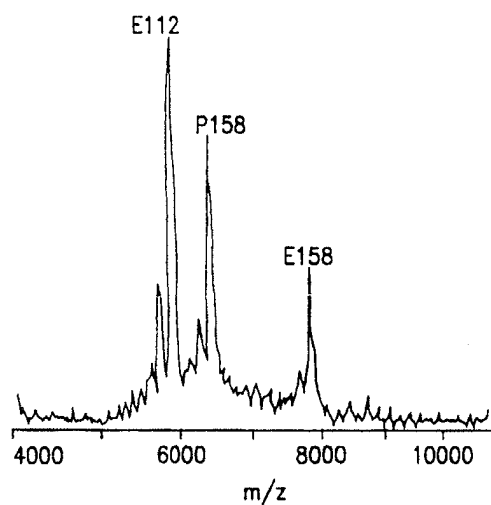
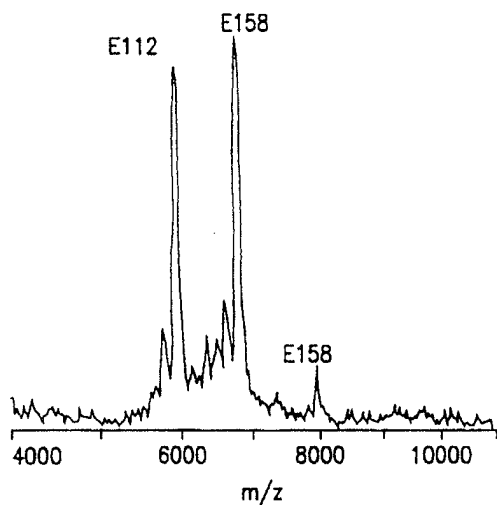


FIG. 60

5'-----GGCGCGACATGGAGGACGTGTGCCGCCGCCTGGT-----
 ε4 ↗ C
 CODON 112
 ↖ ε2 T
 -----TCCGCGATGCCGATGACCTGCAGAAGCGCCTGGC-----3'
 CODON 158



79/123

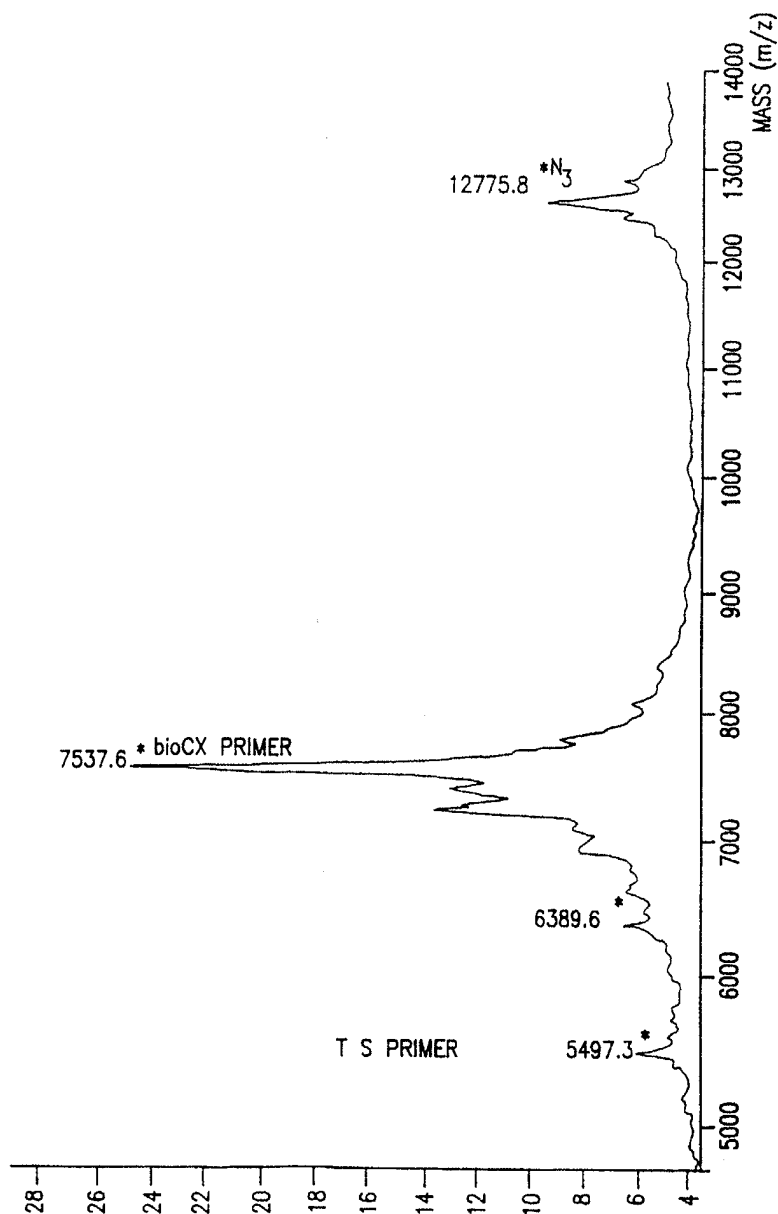


FIG. 62

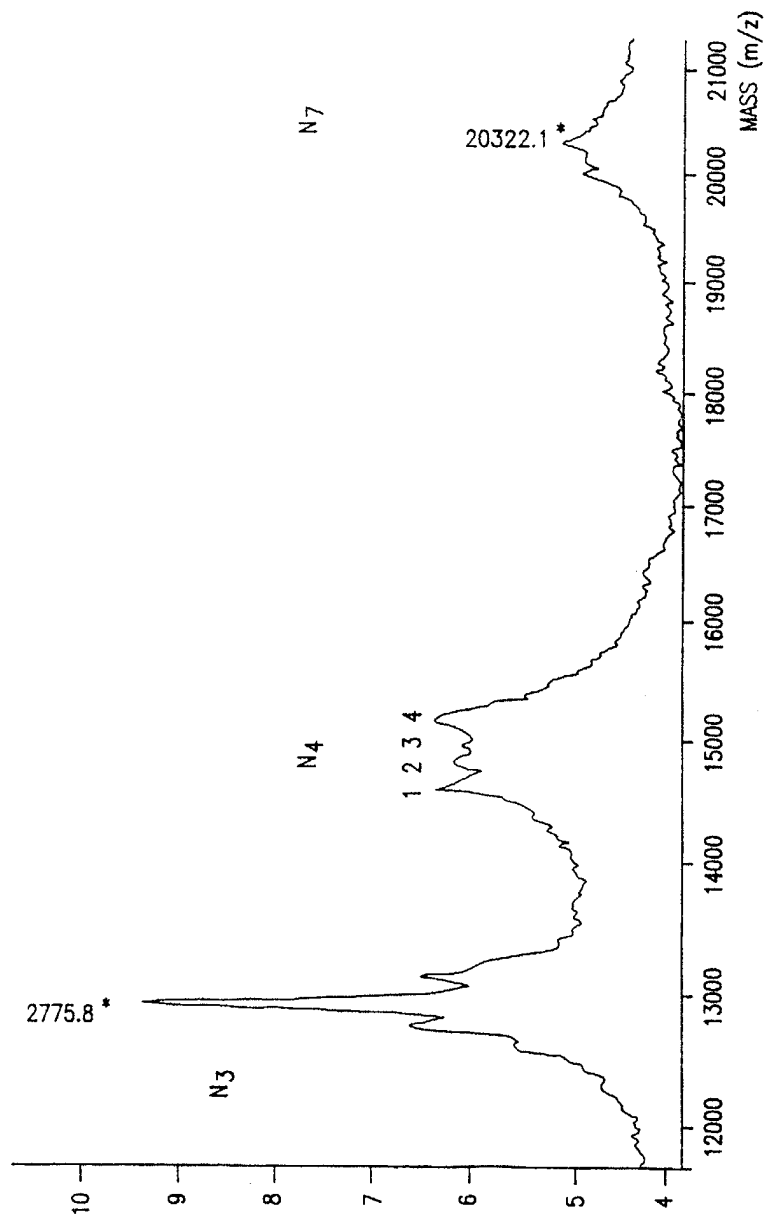


FIG. 63

81/123

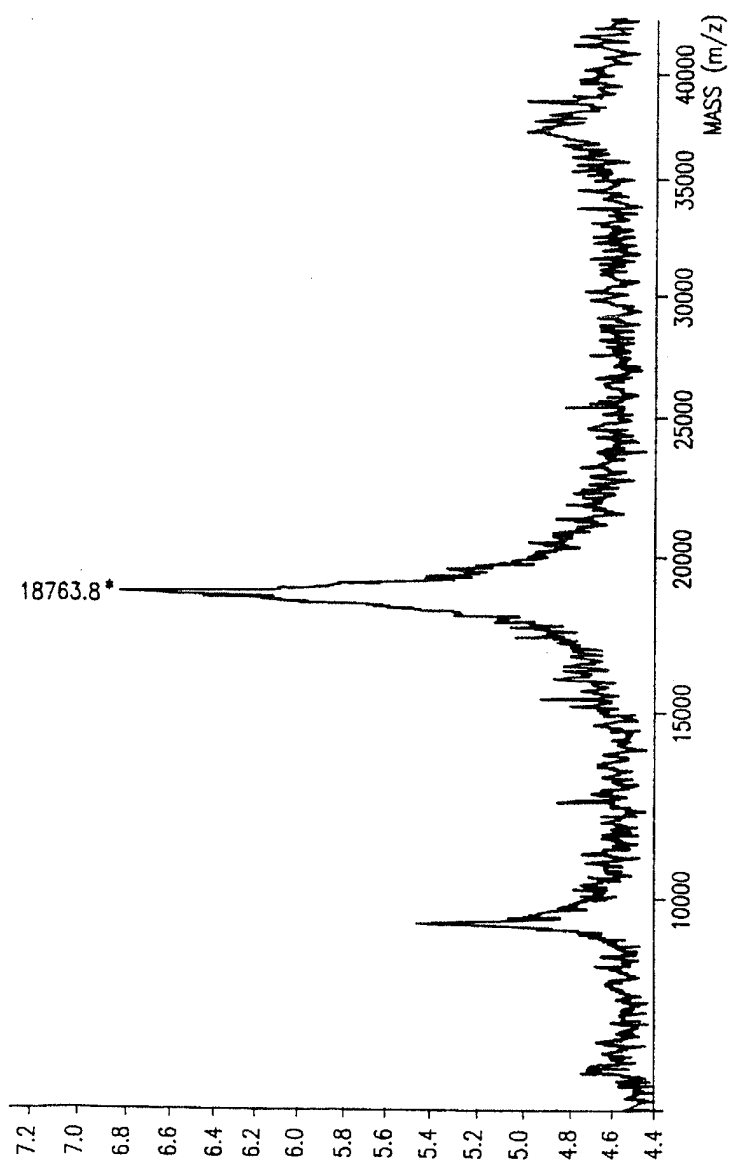


FIG. 64

82/123

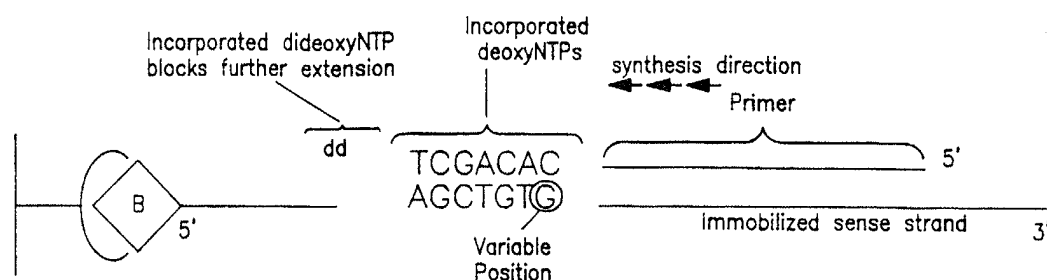


FIG. 65A

ddTTP + dNTP (N=A,C,G)

CGG CTG CGA TCA CCG TGC GG C ACA GCT

WILDTYPE 8246 Da

CGG CTG CGA TCA CCG TGC GG T

6423 Da

CGG CTG CGA TCA CCG TGC GG A ACA GCT

8270 Da

ddATP + dNTP (N=C,T,G)

CGG CTG CGA TCA CCG TGC GG C A

WILDTYPE 6721 Da

CGG CTG CGA TCA CCG TGC GG T A

6736 Da

CGG CTG CGA TCA CCG TGC GG A

6432 Da

FIG. 65B

83/123

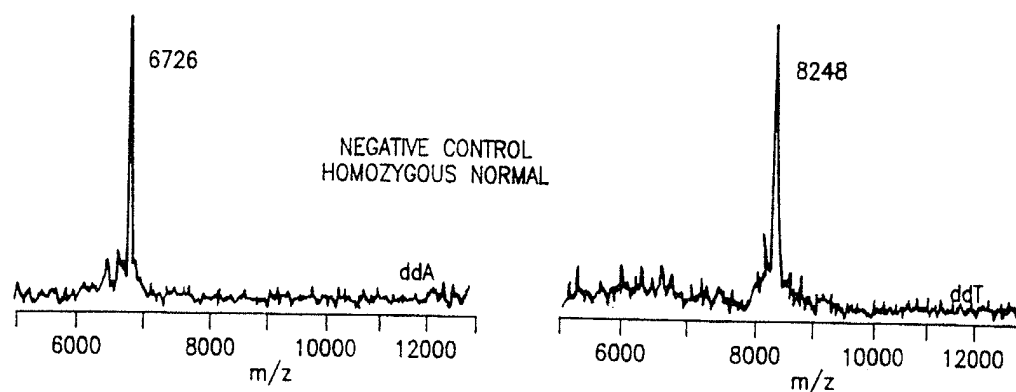


FIG. 66A

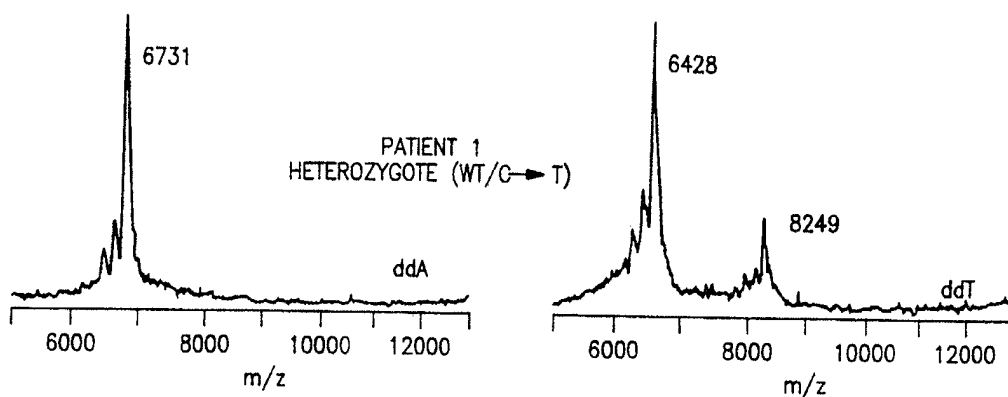


FIG. 66B

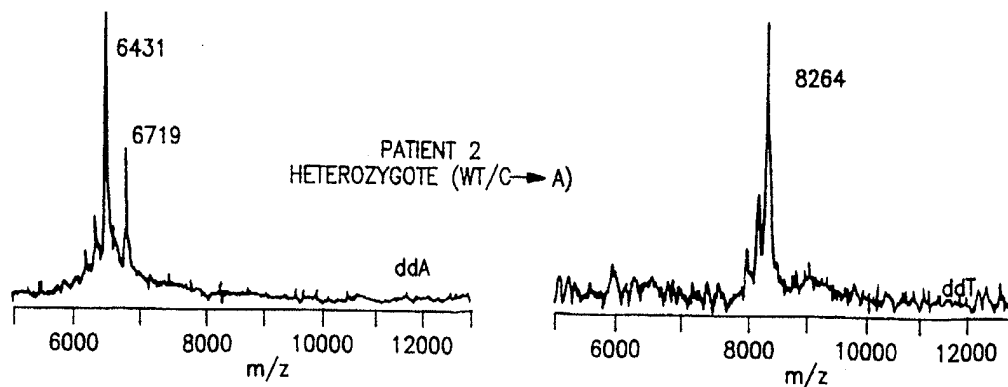
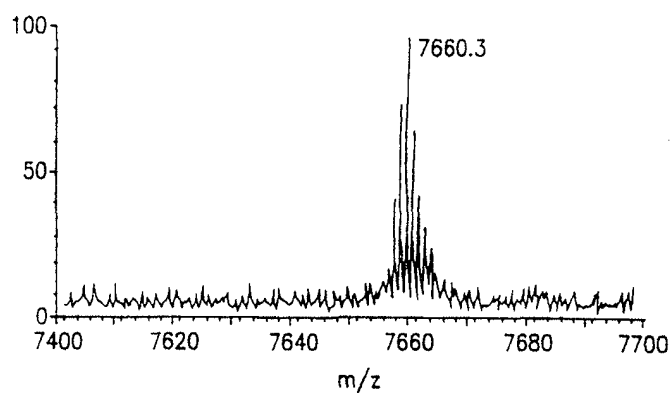
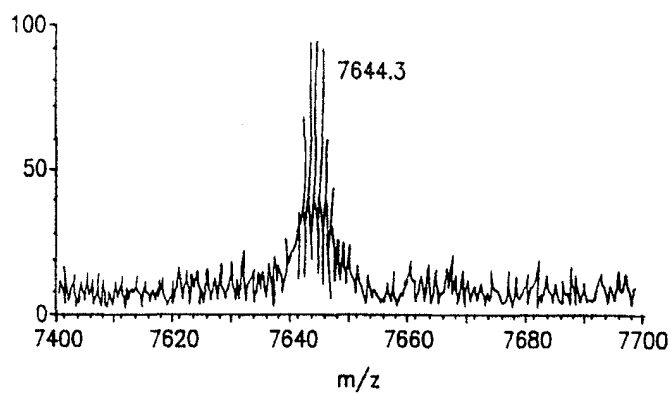
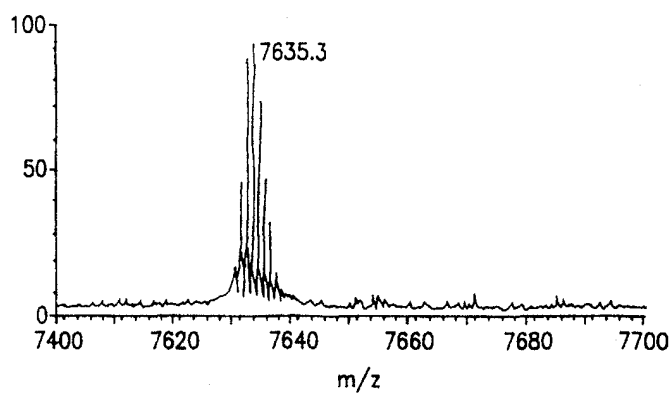


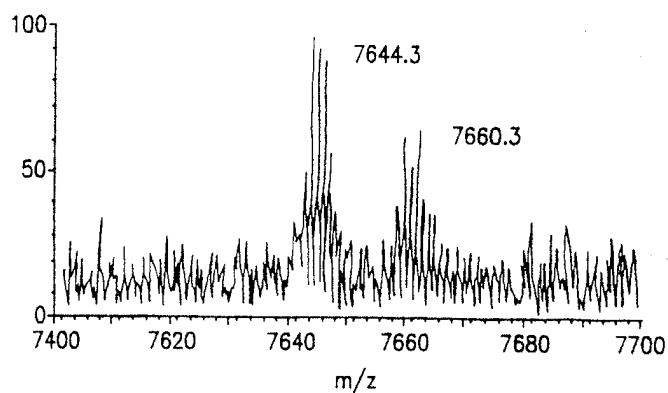
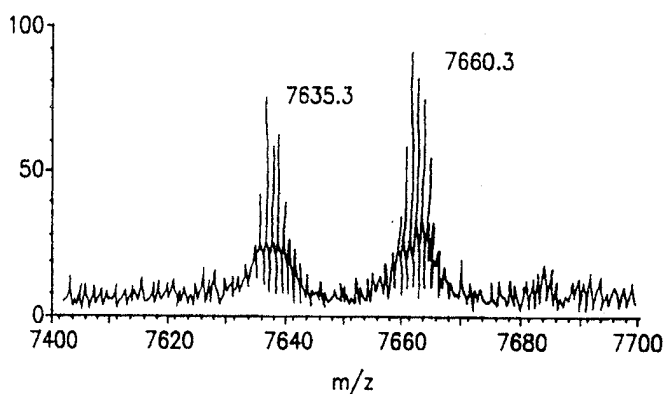
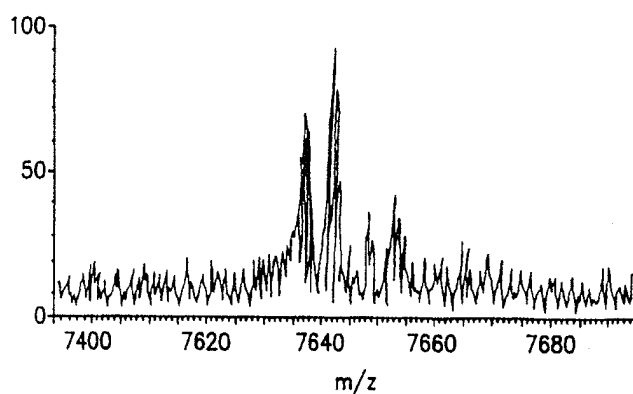
FIG. 66C

SUBSTITUTE SHEET (RULE 26)

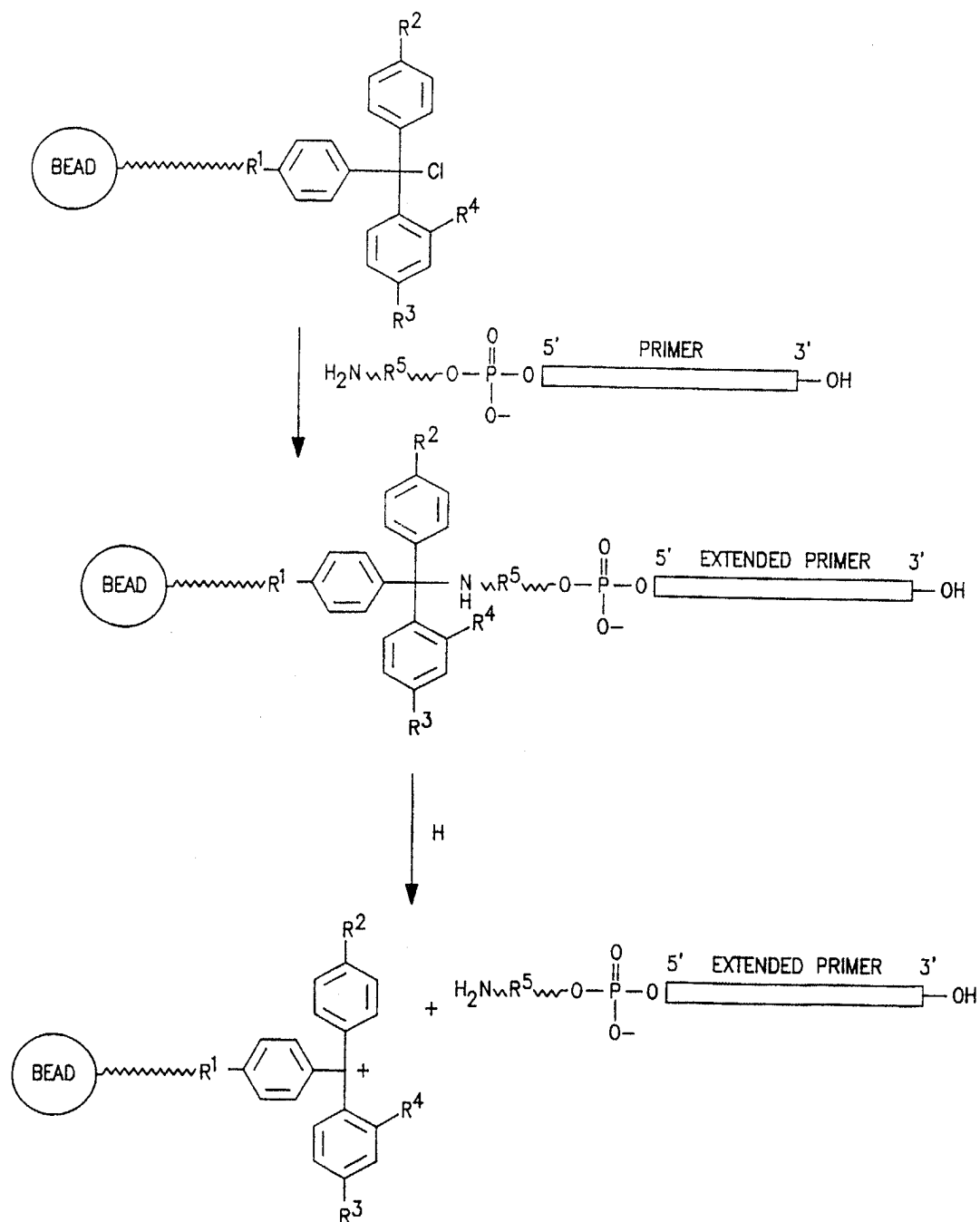
84 / 123

FIG. 67A*FIG. 67B**FIG. 67C*

85/123

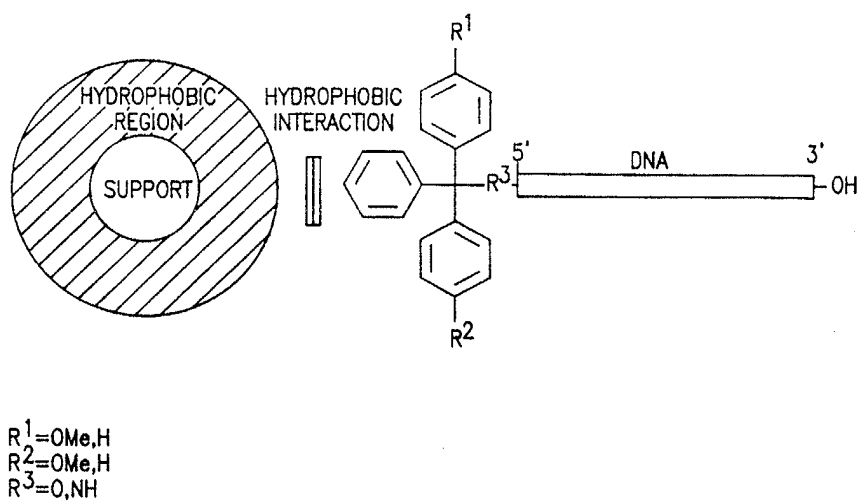
FIG. 67D*FIG. 67E**FIG. 67F*

86 / 123

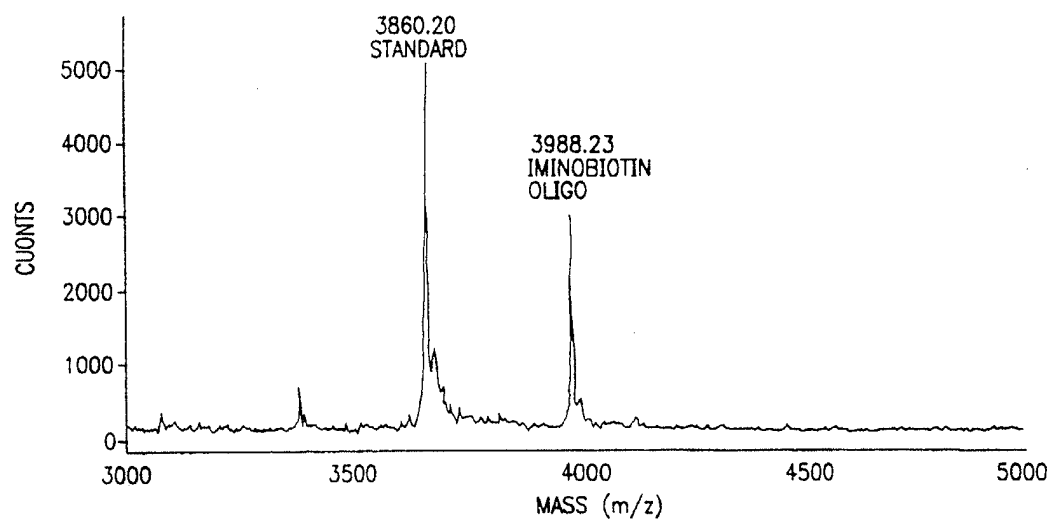
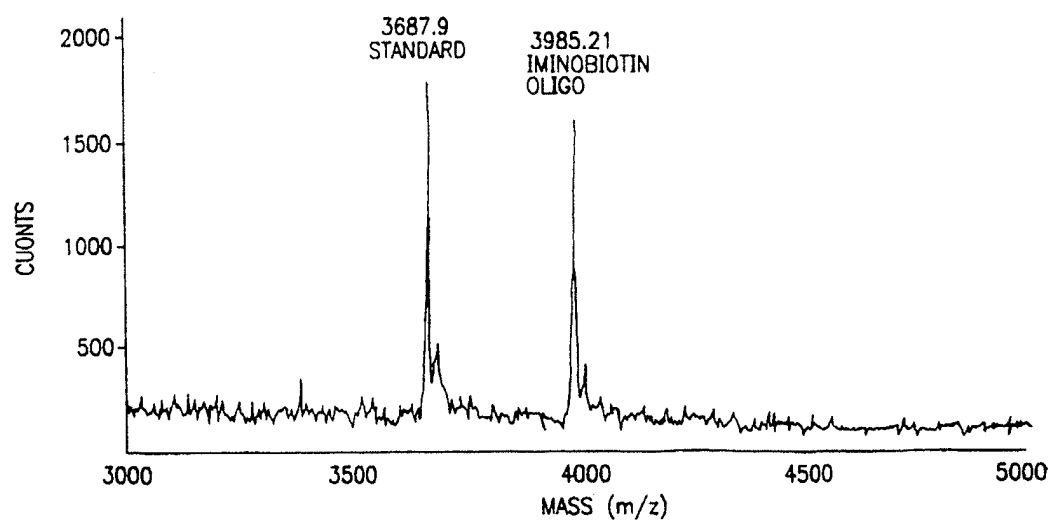


$R^1 = \text{COO}; (\text{CH}_2)_n$; (PARA OR META)
 $R^2 = \text{MeO}; \text{H}$
 $R^3 = \text{MeO}; \text{H}$
 $R^4 = \text{Cl}; \text{H}$
 $R^5 = (\text{CH}_2)_n; (\text{CH}_2)_n \text{CONH}(\text{CH}_2)_n$

FIG. 68

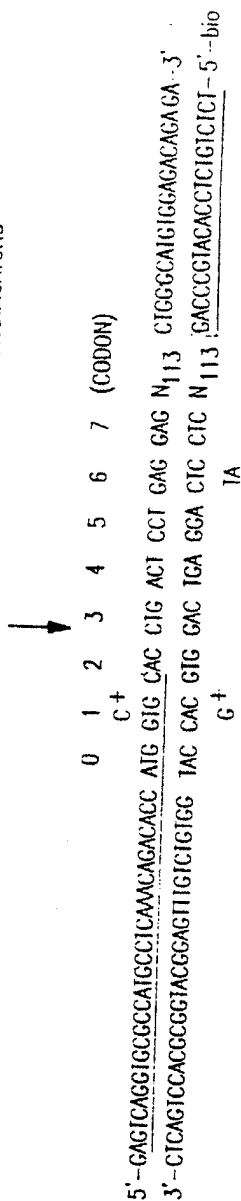
*FIG. 69*

88/123

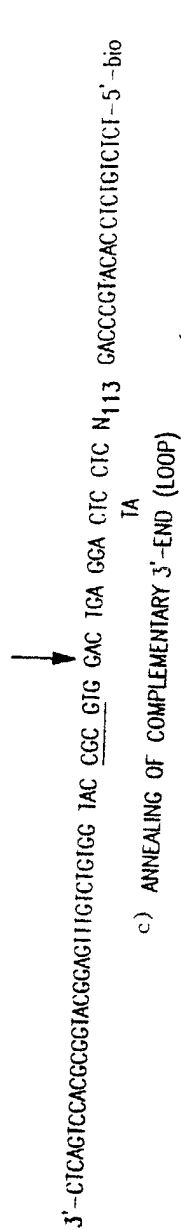
*FIG. 70**FIG. 71*

89/123

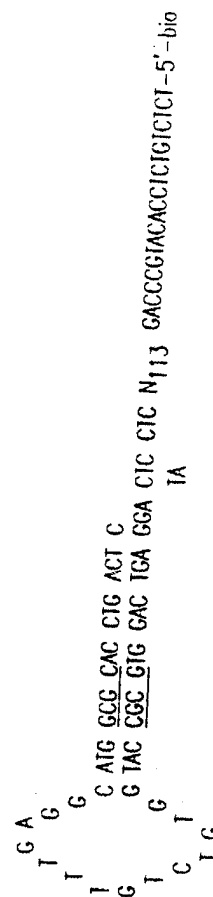
FIG. 72A

LOOP-PROBE FOR MUTATION DETECTION IN THE β -GLOBIN GENE AT CODON 5 AND CODON 6a) PCR-AMPLIFICATION USING β -GLOBIN GENE SPECIFIC PRIMERS WITH MODIFICATIONS

b) DENATURATION OF BIOTIN-STREPTAVIDIN CAPTURED dsDNA



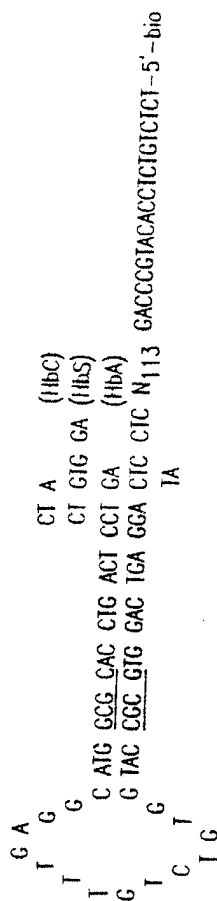
c) ANNEALING OF COMPLEMENTARY 3'-END (LOOP)



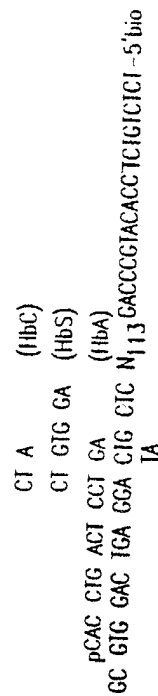
90/123

FIG. 72B

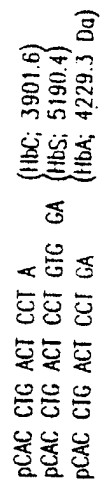
d) PRIMER OLIGO BASE EXTENSION (PROBE) USING ddaIP AND dCTP, dGTP, dTTP



e) Cfo I RESTRICTION ENZYME DIGEST



f) MALDI TOF MASS SPEC ANALYSIS



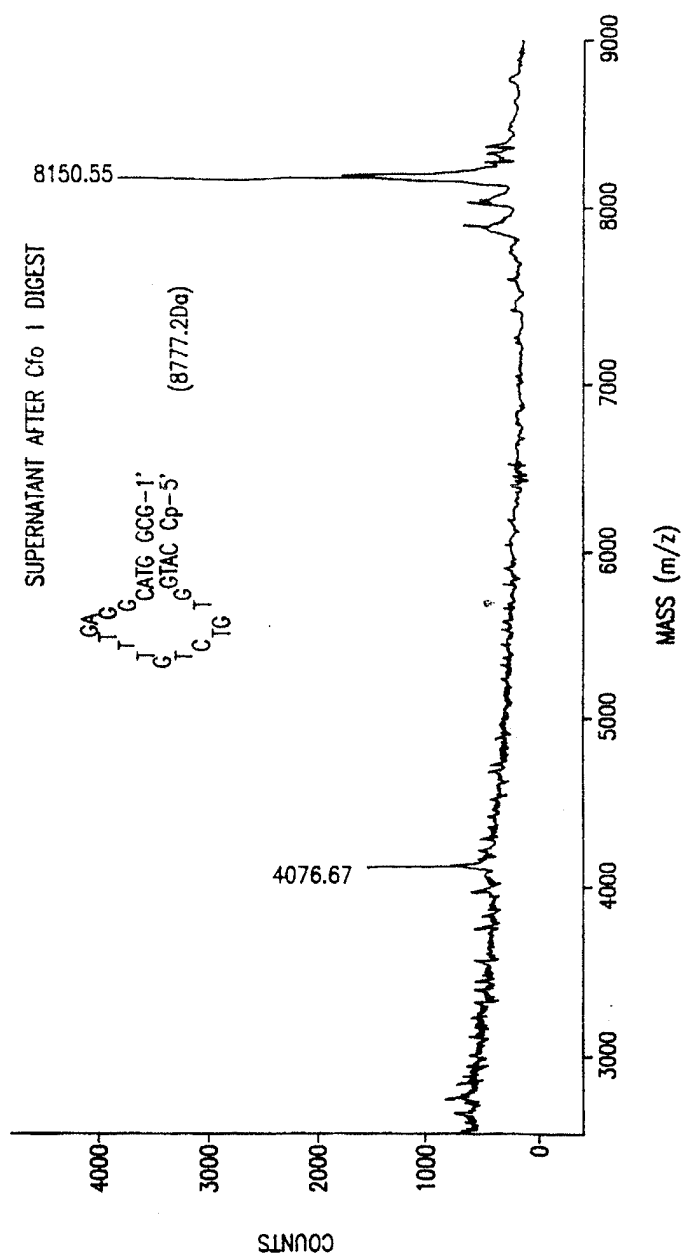


FIG. 73A

92/123

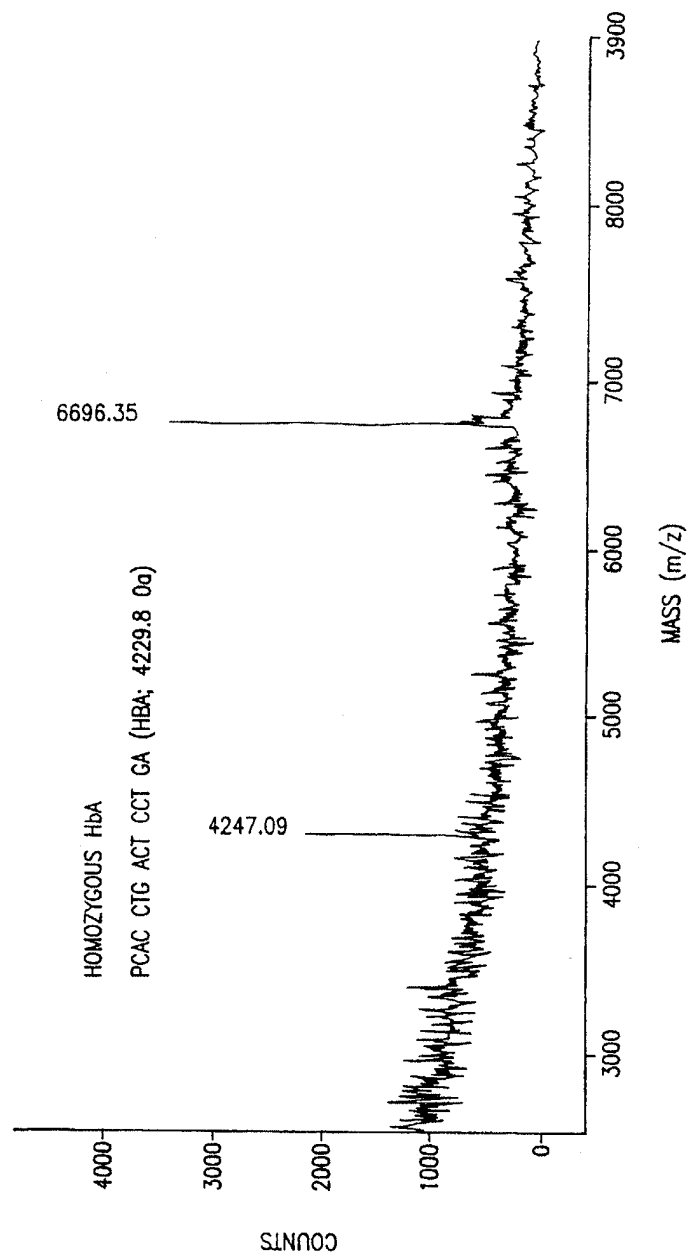


FIG. 73B

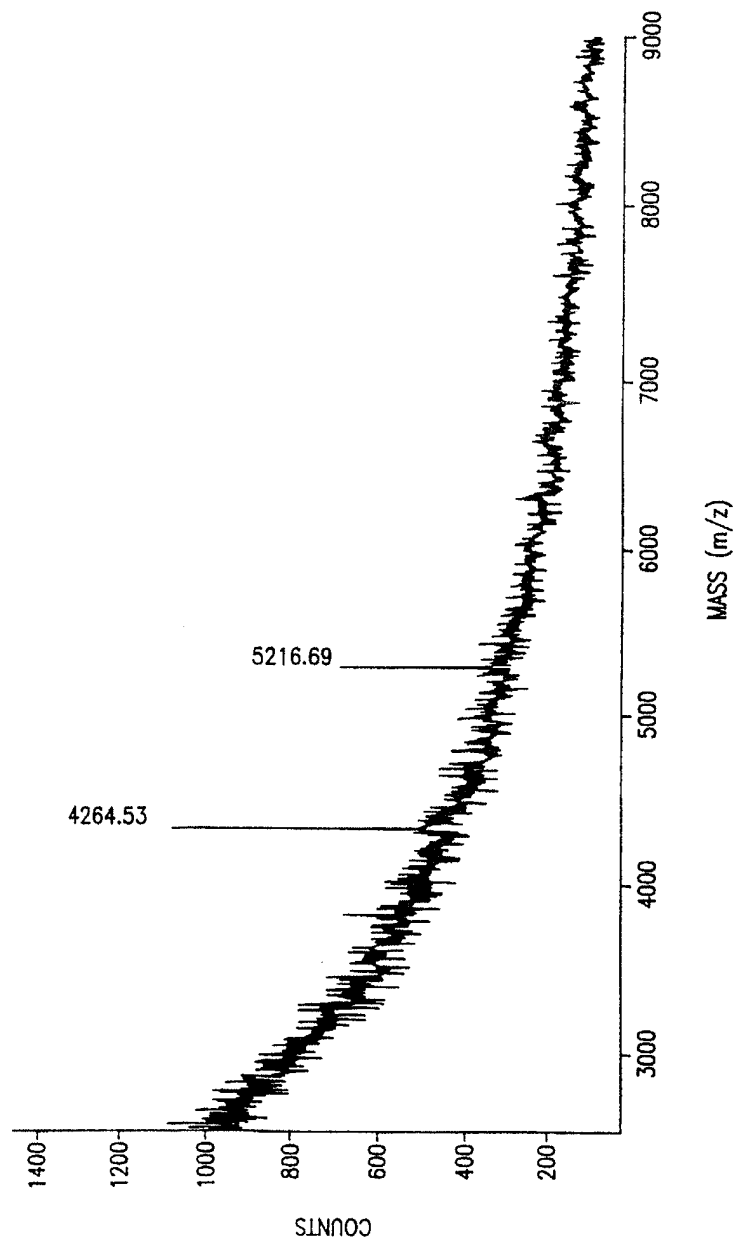


FIG. 73C

94/123

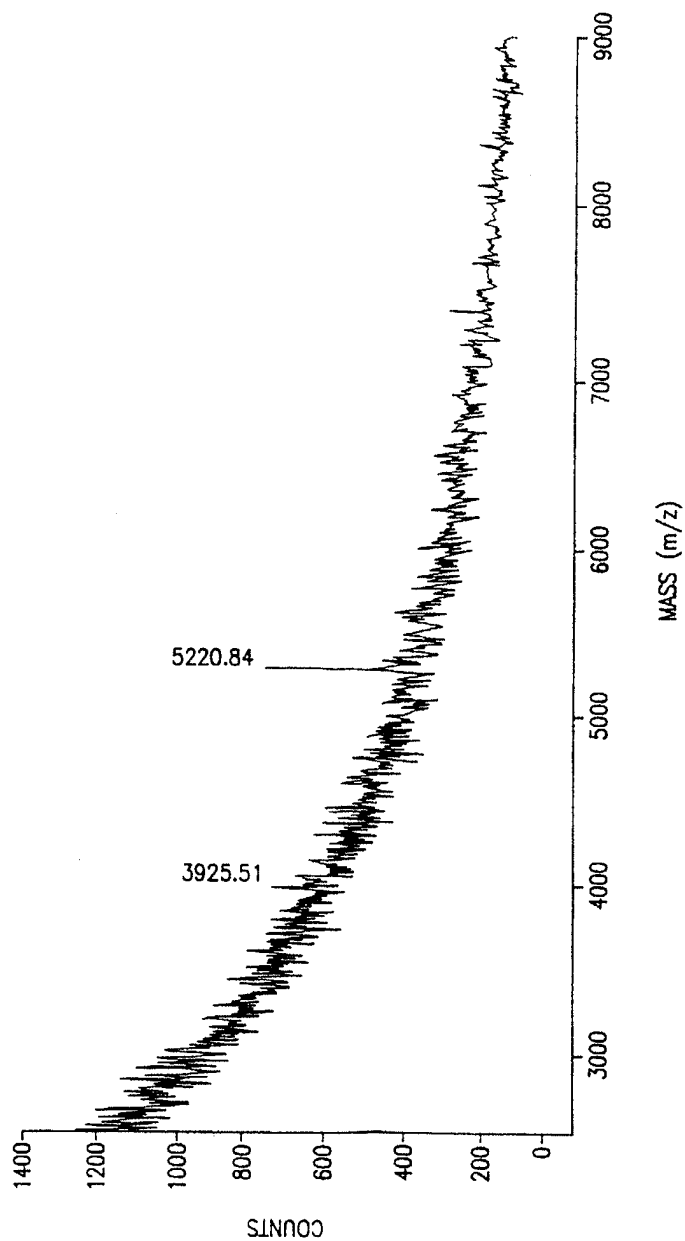


FIG. 73D

95/123

5' CAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTCCAGACATTA
GATAGTCATCTTGGGGCT 3'

FIG. 74

5' ACCTAGCGTTCAGTTCGACTGAGATAATACGACTCATATAGCAGCTCTCATTTTCCATAC3'
RANDOM SEQUENCE T7 PROMOTER SEQUENCE CKR-5 HOMOLOGUE

FIG. 75

96/123

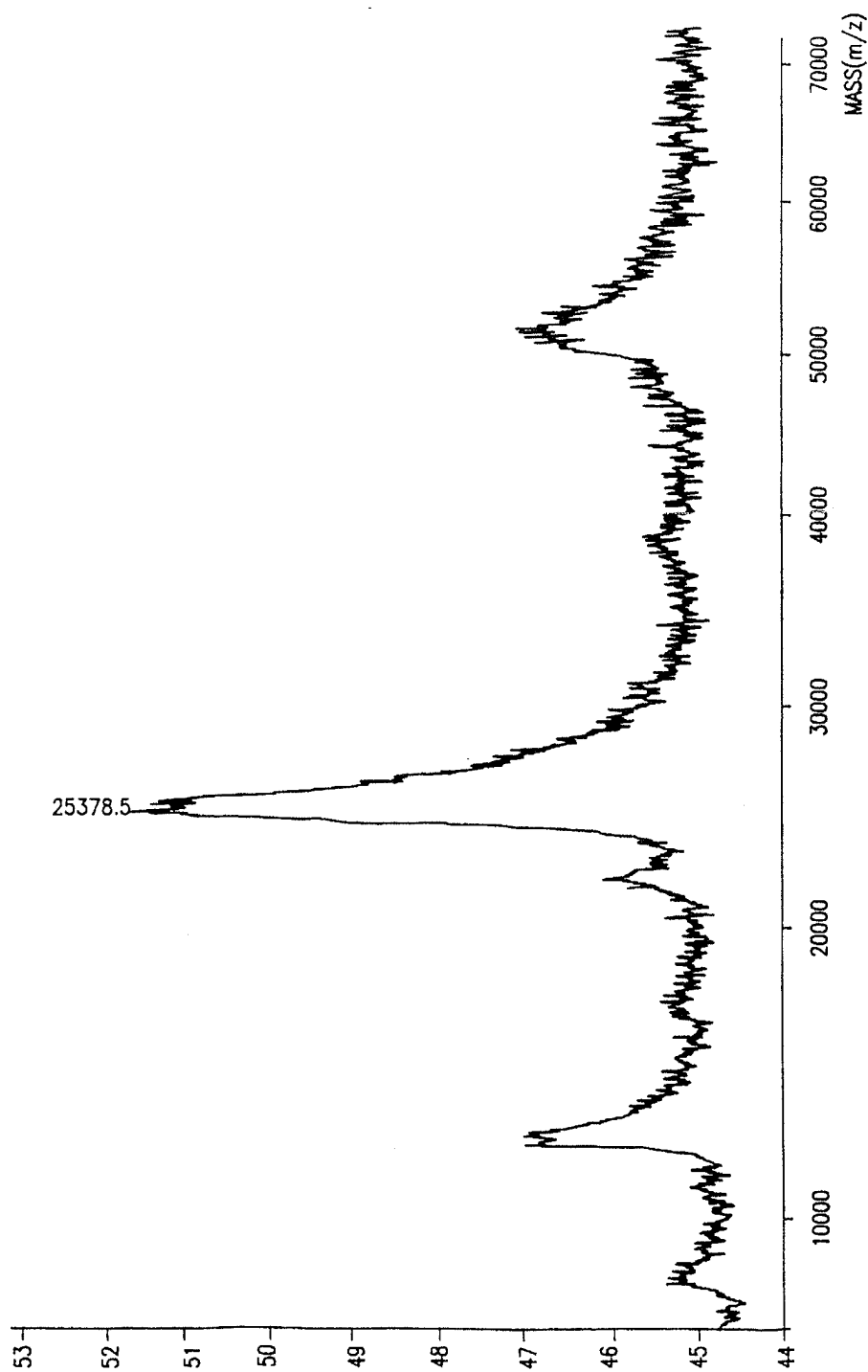


FIG. 76

97/123

FIG. 77A

RNase T₁
GpN →

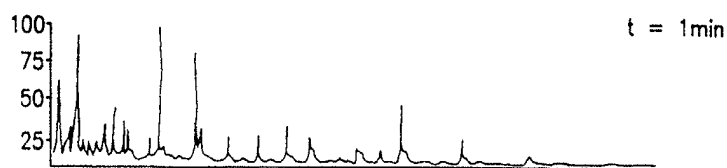


FIG. 77B

ALKALINE
HYDROLYSIS

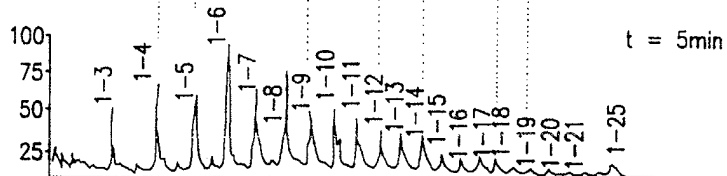


FIG. 77C

U₂
ApN →

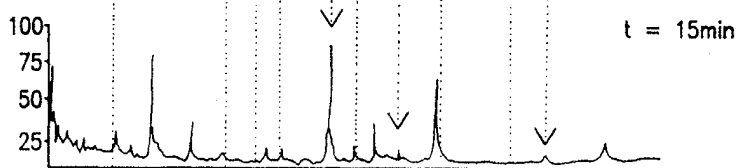


FIG. 77D

PhyM
ApN →
UpN →

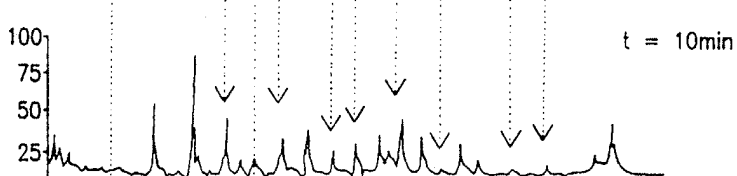
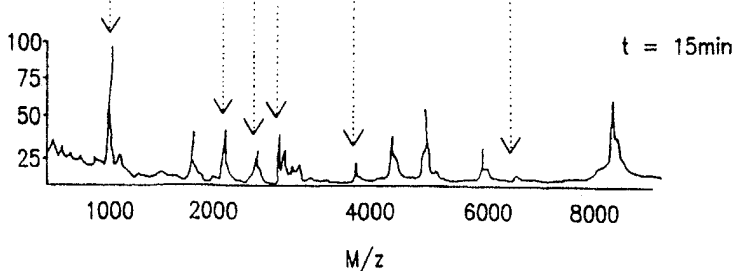


FIG. 77E

A
UpN →
GpN →



98/123

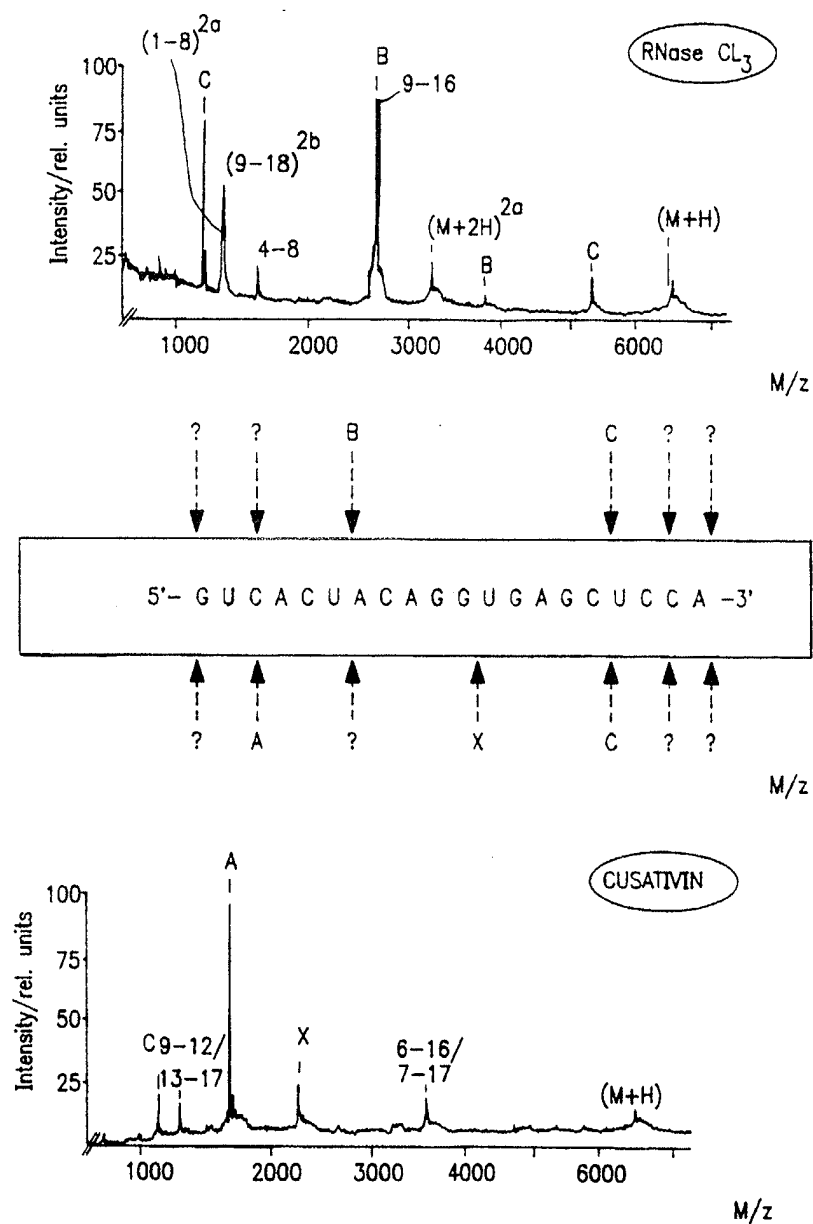


FIG. 78

99/123

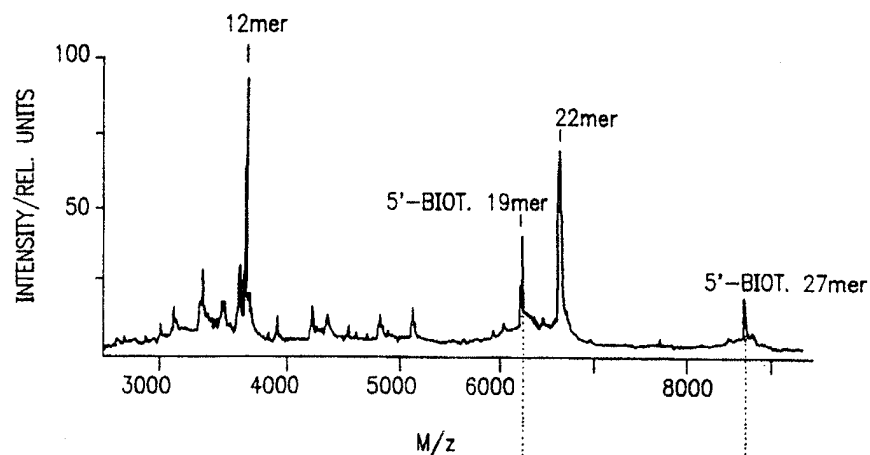


FIG. 79A

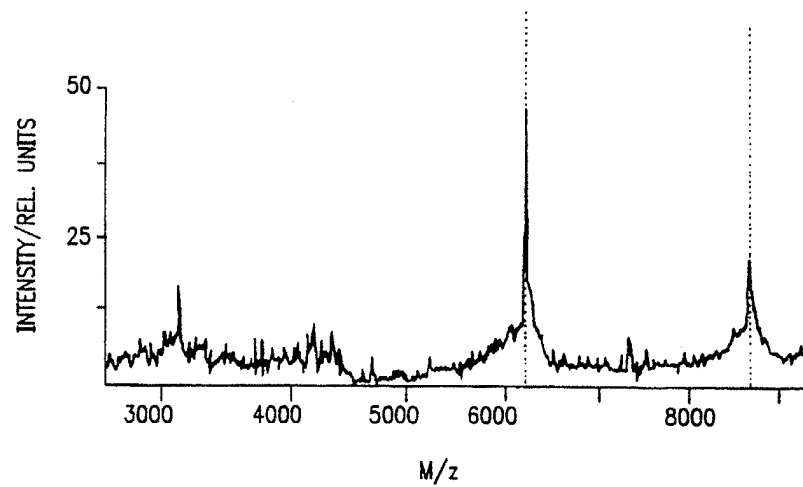


FIG. 79B

100/123

FIG. 80A

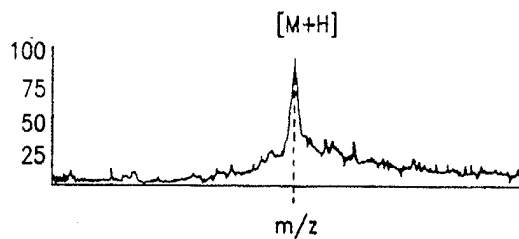


FIG. 80B

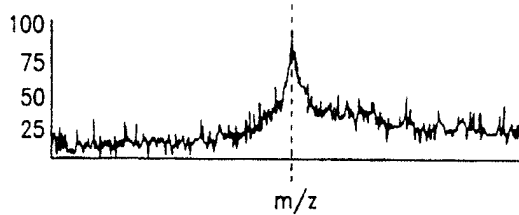


FIG. 80C

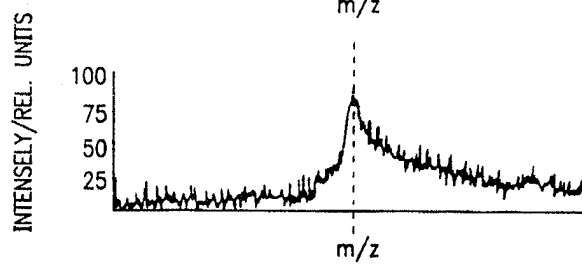


FIG. 80D

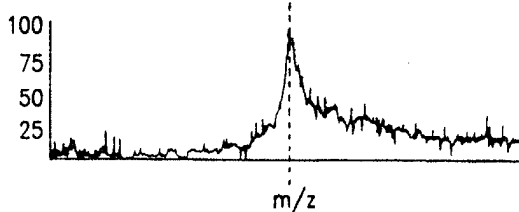
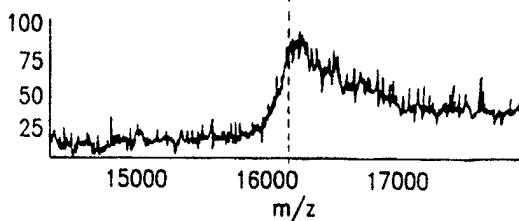


FIG. 80E



101/123

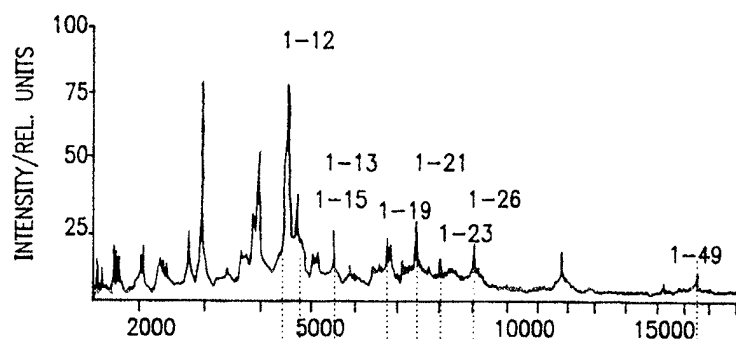


FIG. 81A

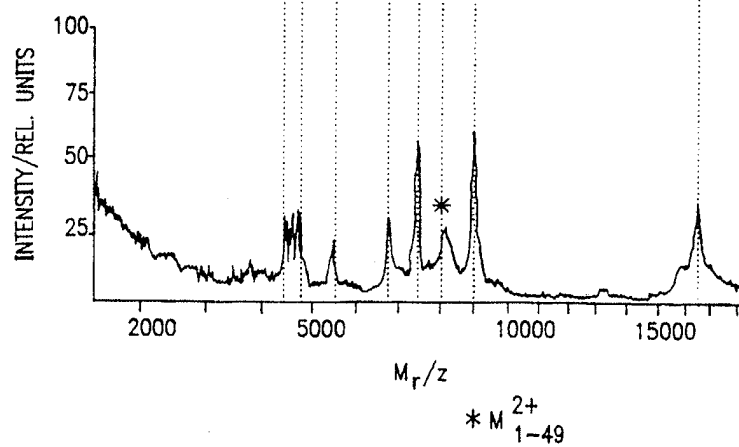


FIG. 81B

102 / 123

FIG. 82A

β -TAG1
 GTGTCCTC TCACACAACCT GTGTCACTA GCAACCTCAA ACACACACC ATG GTG CAC CTG ACT C
 1 2 3 4 5 6 7 8 (CODON)
 GTGTCCTC ATG GTG CAC CTG ACT C
 -CT GAG GAG AAG TCT GCC GTT
 THALASSEMIA (COD5 ACT)
 A HbC ---
 T HbS

β -TAG2
 CCCTGTGGT GAG GCC CTG GGC A
 27 28 29 30 (CODON)
 ACT GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT GGT GAG GCC CTG GGC AG GTTGGTATCA AGGTACAAG
 A THALASSEMIA (IVS -1 nt1 G>A)
 C THALASSEMIA (IVS -1 nt5 G>C)
 CTHALASSEMIA (IVS -1 nt6 T>C)

ACAGGTTAA GCACACCAAT AGAACTGGG CATGTGGACA CAGACAAG
 β -11

FIG. 82B

(1) SINGLE STRANDED
 (STV) BIO3' - ACCTCAGTCCATCGTCAGGACGAG - 5'
 UNIVERSAL SS SEQUENCE | COMPLEMENTARY
 E.G. β -TAG1

(2) PARTLY DOUBLE STRANDED
 5'-TGCAGTCAGGTACGAGTC-3'
 (STV) BIO 3'-ACCTCAGTCCATCGTCAGGACGAG - 5'
 UNIVERSAL DS SEQUENCE | COMPLEMENTARY
 E.G. β -TAG2

103/123

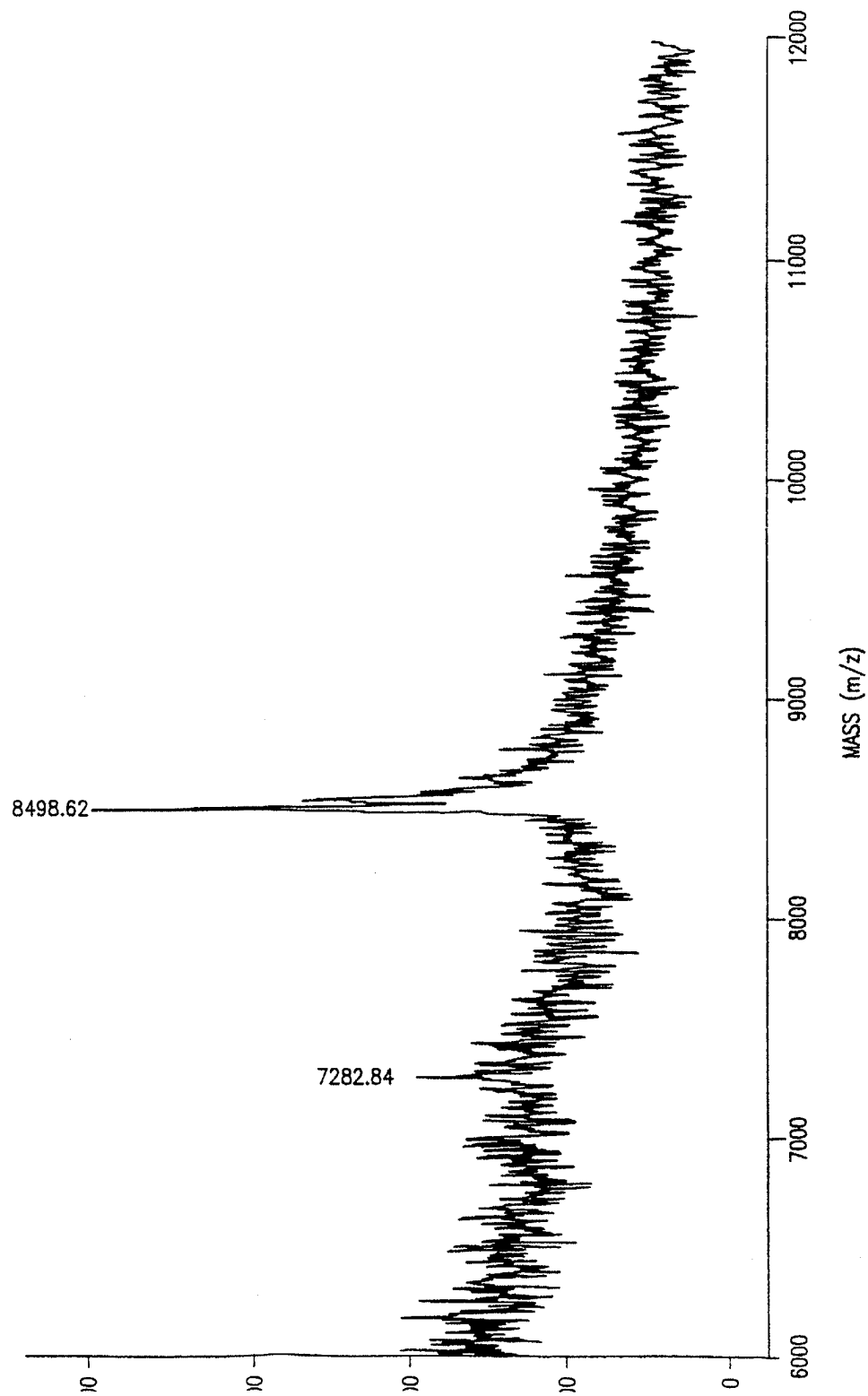


FIG. 83

104/123

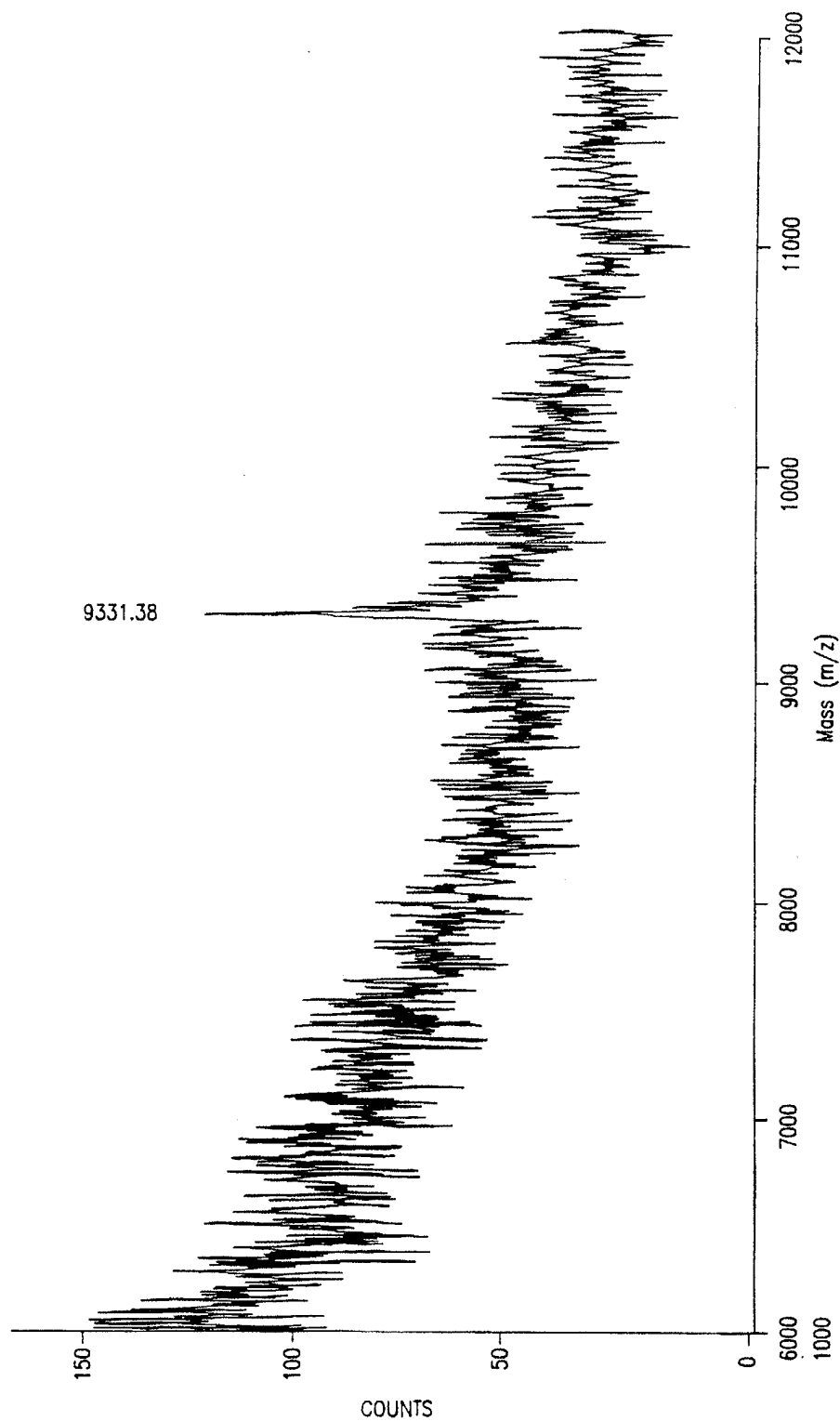
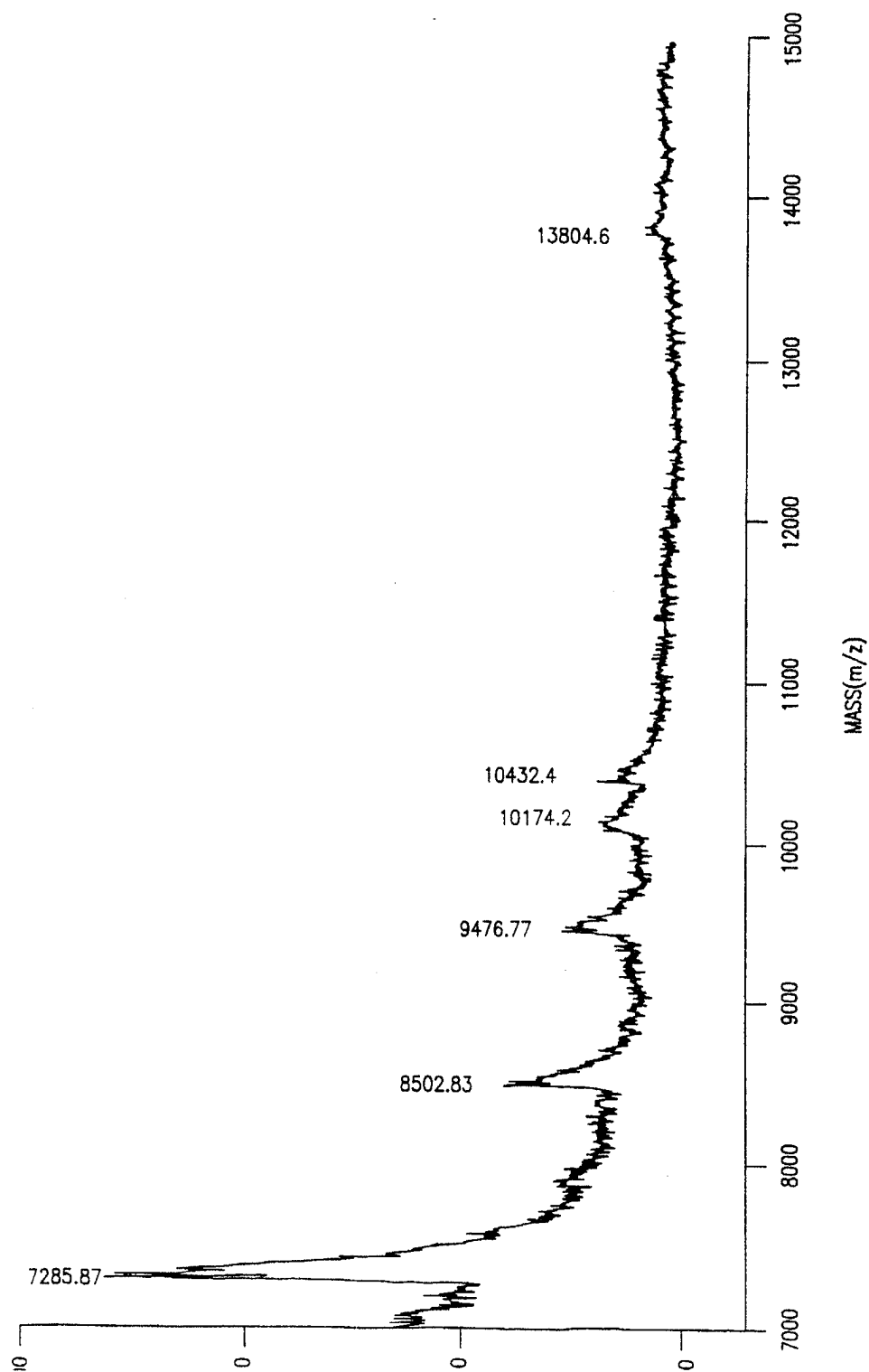


FIG. 84

105/123



MASS(m/z)

FIG. 85

106 / 123

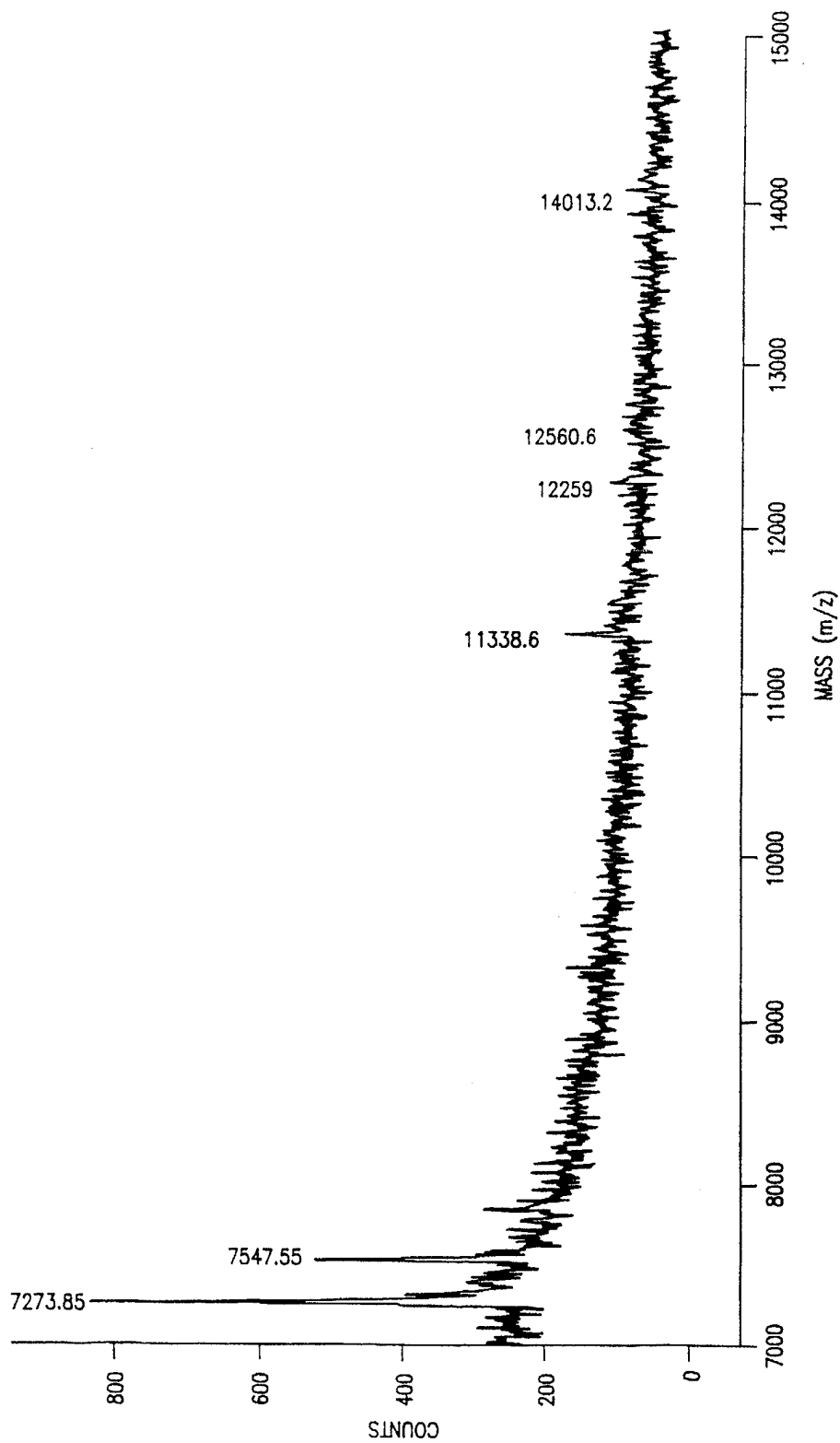


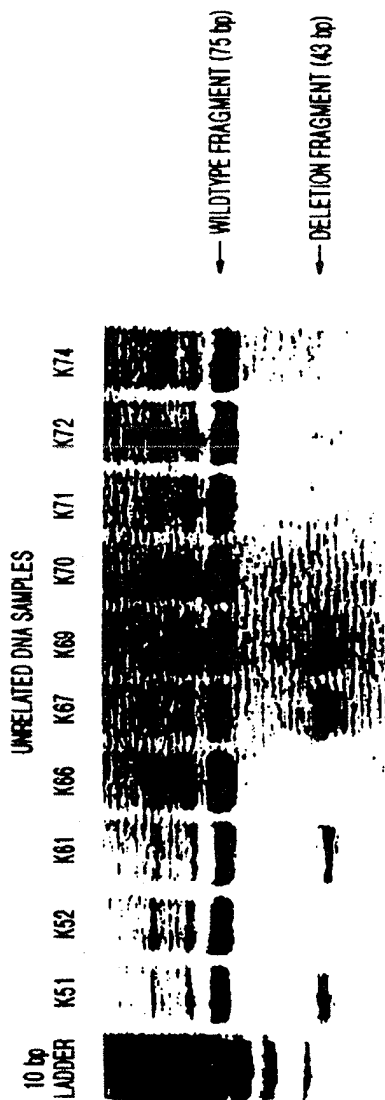
FIG. 86

107/123

<i>FIG. 87A</i>	PCR PRODUCT CKR _Δ F CAGCTCTCAT TTTCATACAT GTCAGTATCA ATTCTGGAAG AATTTCAGACA TTAAAGAT AGTCATCTTG GGGCT GTCGAGAGTA AAGGTATGT CAGTCATAGT TAAGACCTTC TTAAAGGTCT CKR _Δ R-BIO GTAAATTCTA TCAGTAGAAC CCCGA-BIOTIN	
<i>FIG. 87B</i>	SIZE ANALYSES: WILDTYPE SENSE STRAND W/O A: CAGCTCTCAT TTTCATACAT GTCAGTATCA ATTCTGGAAG AATTTCAGACA CATTAAAGAT AGTCATCTTG GGGCT SENSE STRAND WITH A: CAGCTCTCAT TTTCATACAT GTCAGTATCA ATTCTGGAAG AATTTCAGACA CATTAAAGAT AGTCATCTTG GGGCTA	75 bp 23036 Da 76 bp 23349 Da
<i>FIG. 87C</i>	SIZE ANALYSES: 32 bp DELETION SENSE STRAND W/O A: CAGCTCTCAT TTTCATACAT TTAAGAT AGTCATCTTG GGGCT SENSE STRAND WITH A: CAGCTCTCAT TTTCATACAT TTAAGAT AGTCATCTTG GGGCTA	43 bp 13143 Da 44 bp 13456 Da
<i>FIG. 87D</i>	PROBE ANALYSES: WILDTYPE (ddITP TERMINATION): CAGCTCTCAT TTTCATACAT GT	22 bp 6604 Da
<i>FIG. 87E</i>	PROBE ANALYSES: 32 bp DELETION (ddITP TERMINATION): CAGCTCTCAT TTTCATACAT T	21 bp 6275 Da

108/123

DETECTION OF A 32 bp DELETION IN THE CHEMOKINE RECEPTOR BY PCR AND 15% NATIVE PAGE



DELETION FRAGMENTS ARE ONLY WEEKLY STAINED BY SILVER

FIG. 88

109/123

HETEROZYGOUS

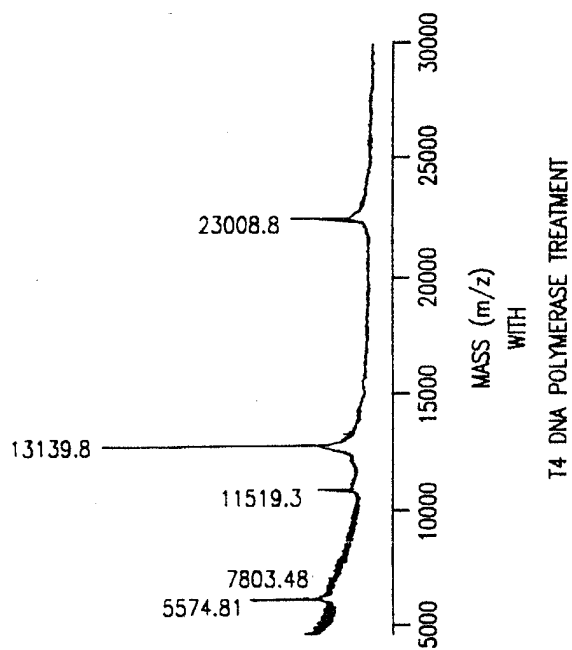


FIG. 89B

HETEROZYGOUS

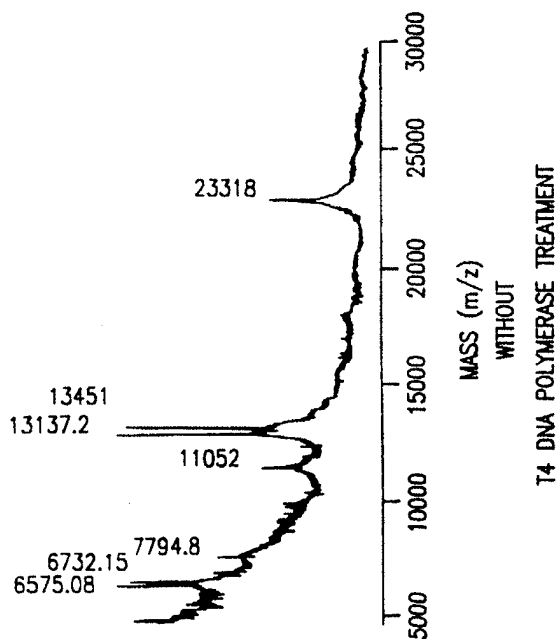


FIG. 89A

110/123

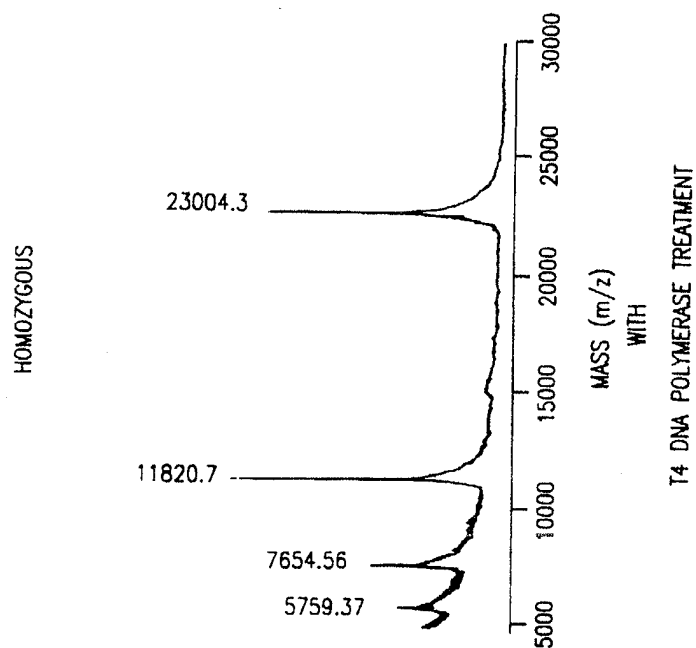


FIG. 89D

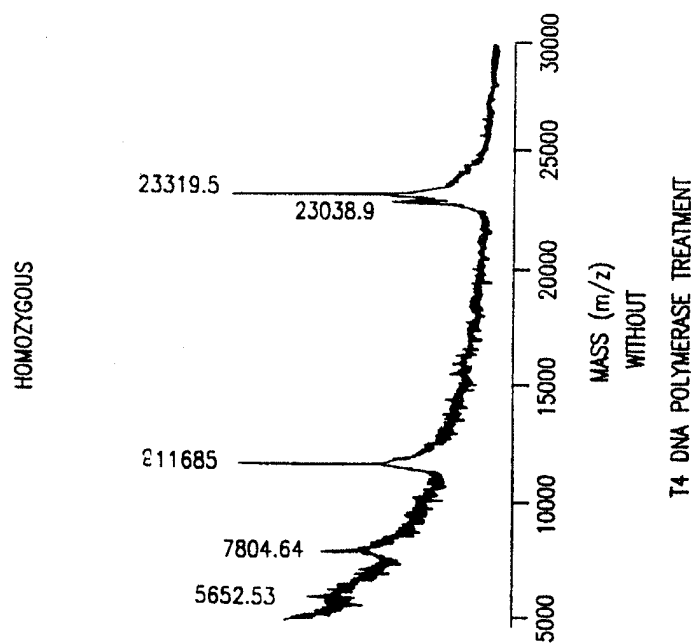


FIG. 89C

111/123

FIG. 90A

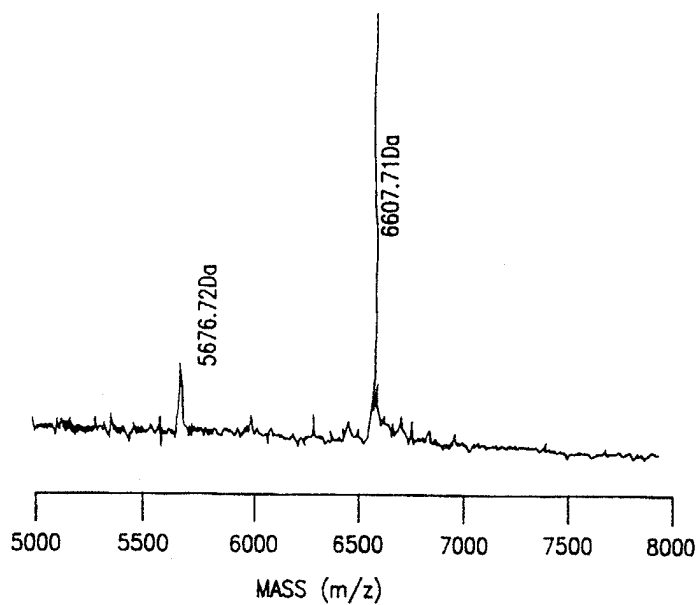
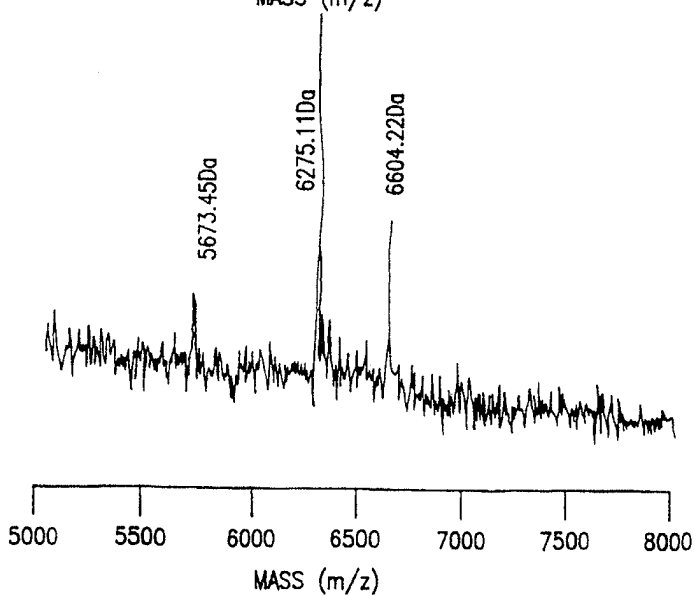


FIG. 90B



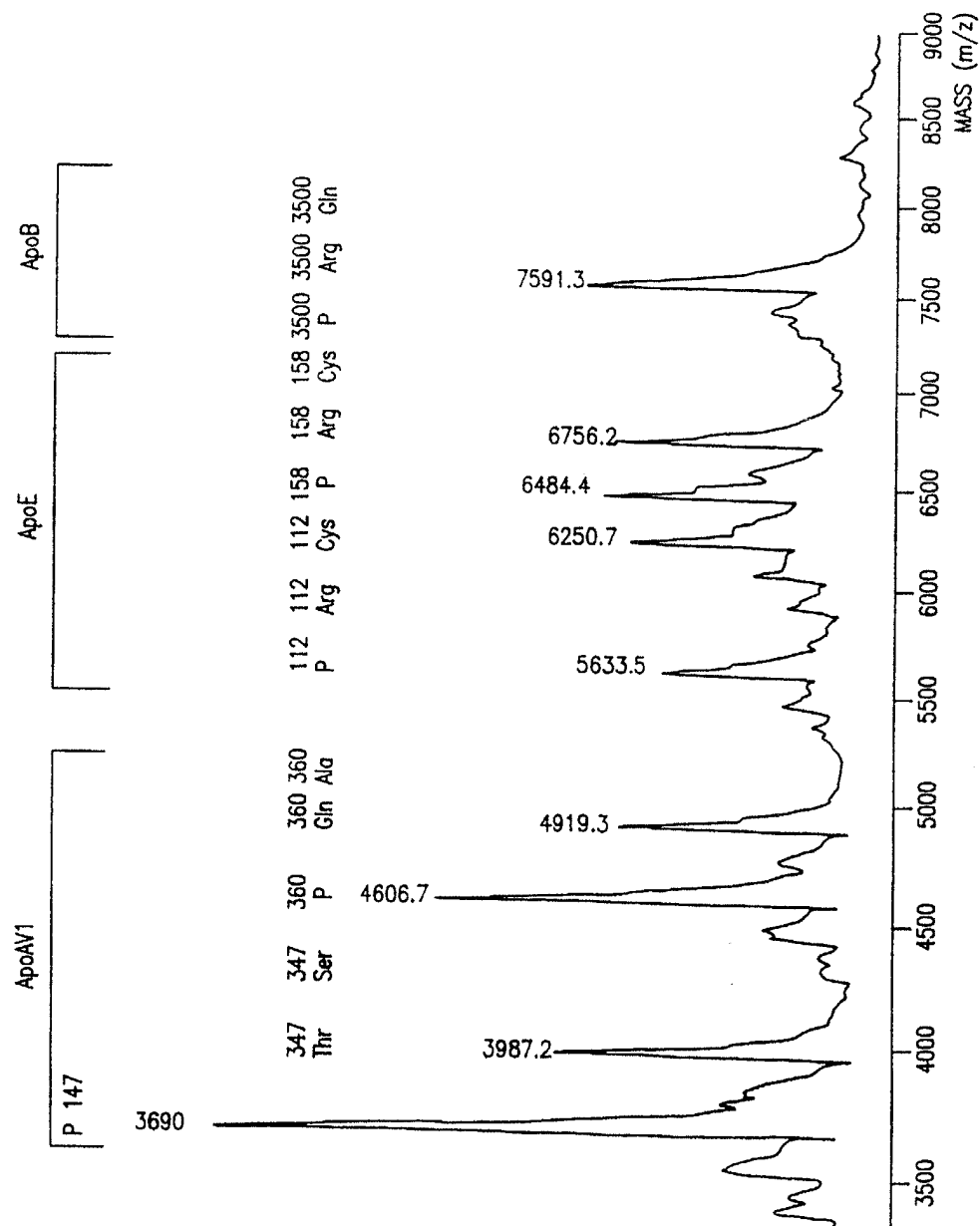


FIG. 91

113/123

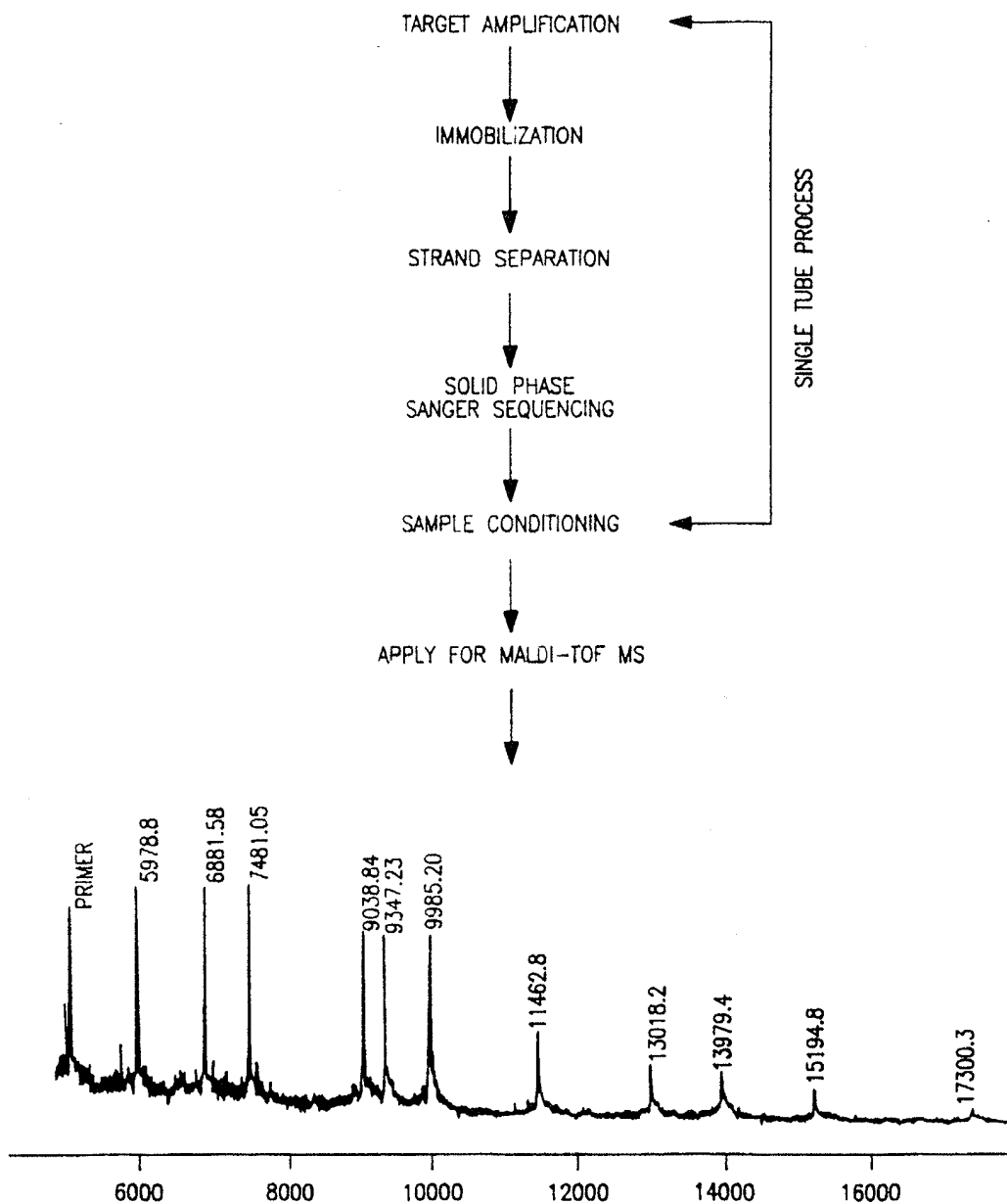


FIG. 92

114/123

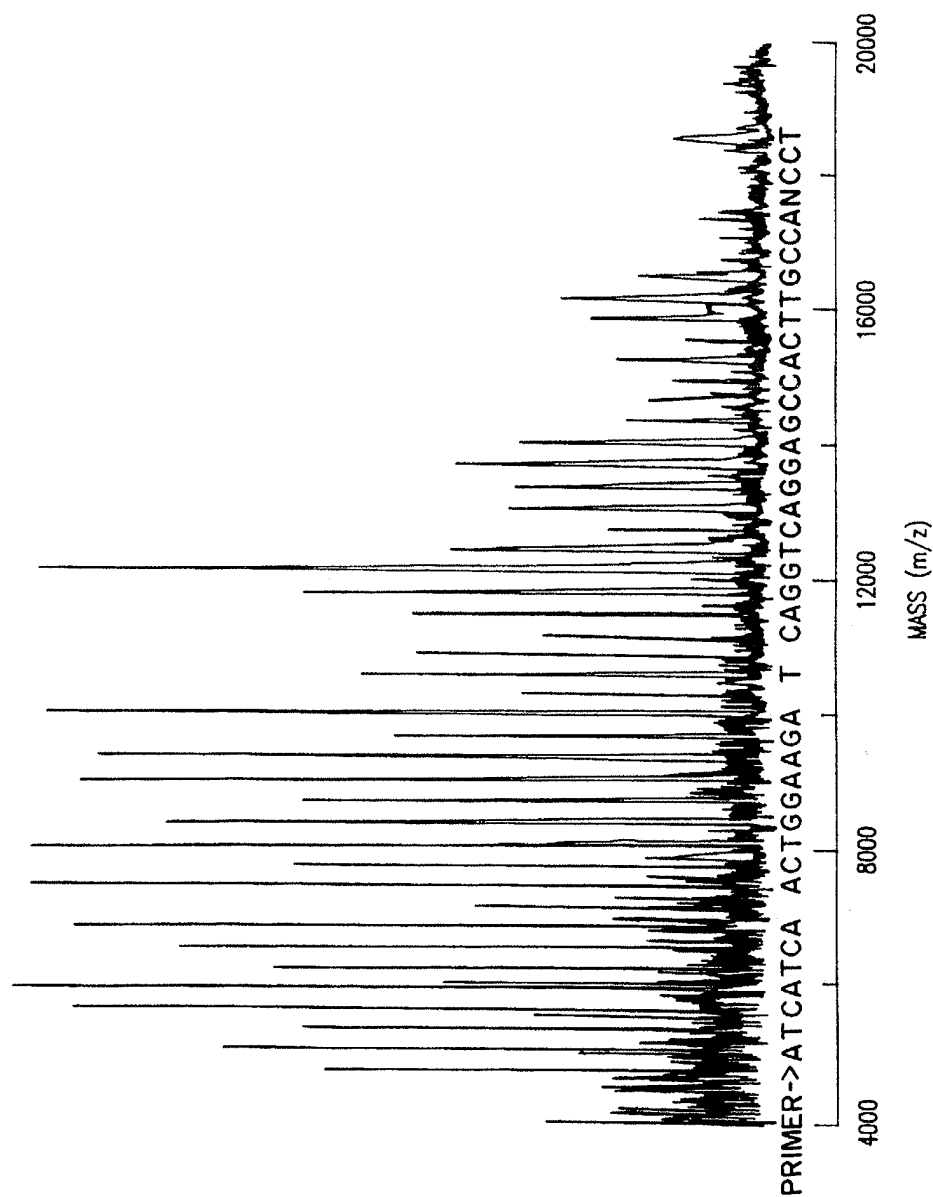
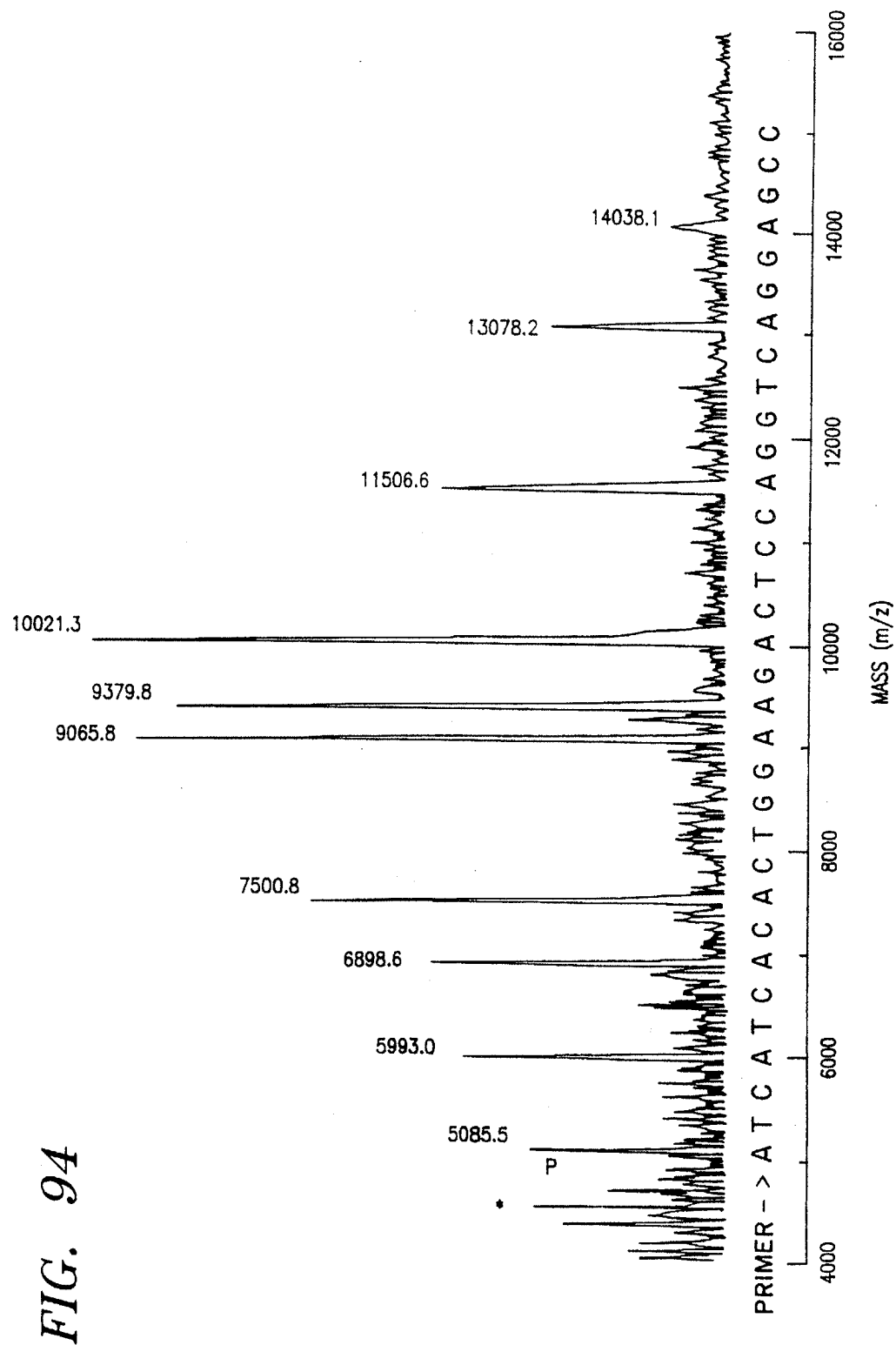


FIG. 93

115/123



116/123

1. MICROITER PLATE ISOTHERMAL SEQUENCING
2. 'PIEZOELECTRIC PIPETTE' TRANSFER
3. CORETECH MS

P-AICCACTACACTACATGCTGTAACAGTTGGwGCwwGddCddC
 5 10 15 20 25 30 35 40

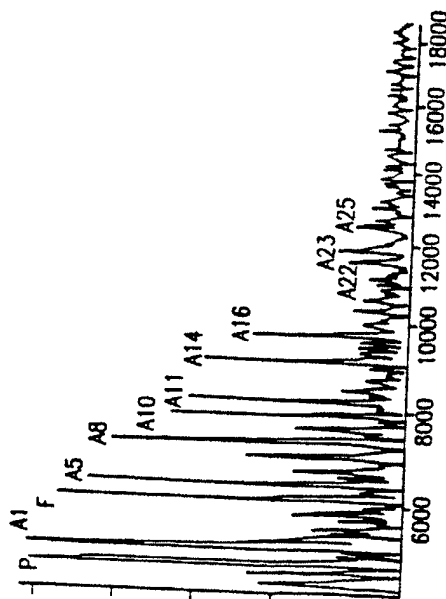


FIG. 95A

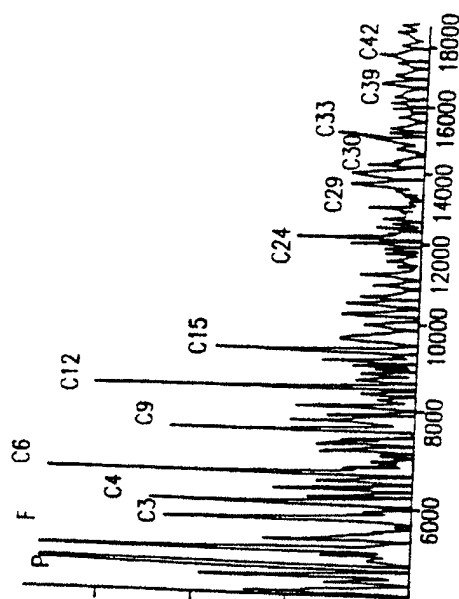


FIG. 95B

117/123

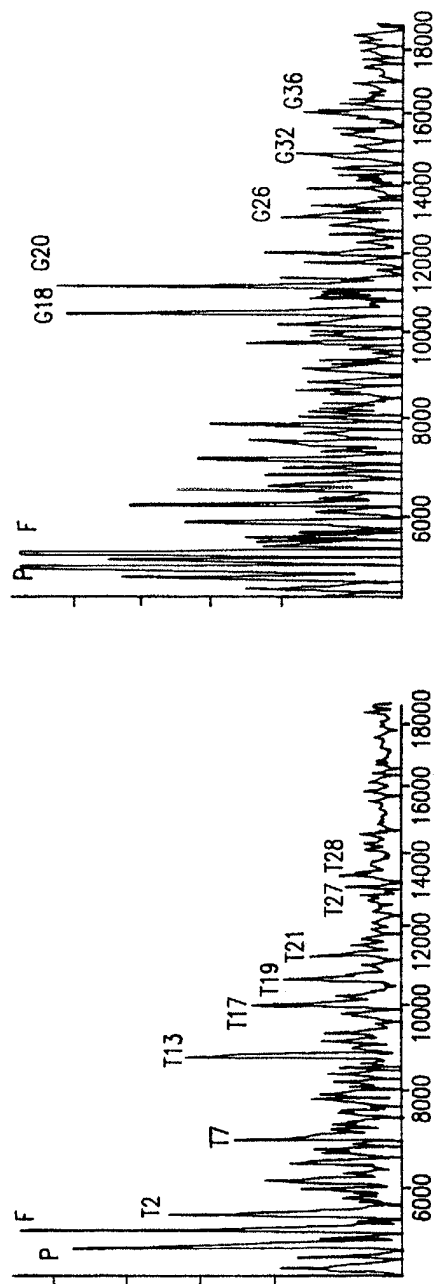


FIG. 95D

FIG. 95C

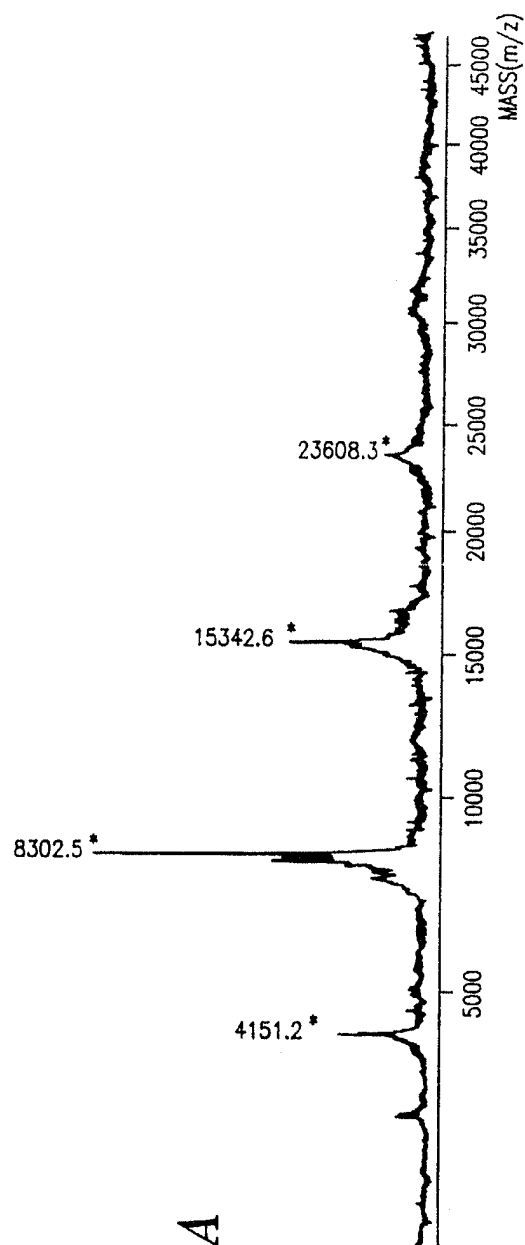


FIG. 96A

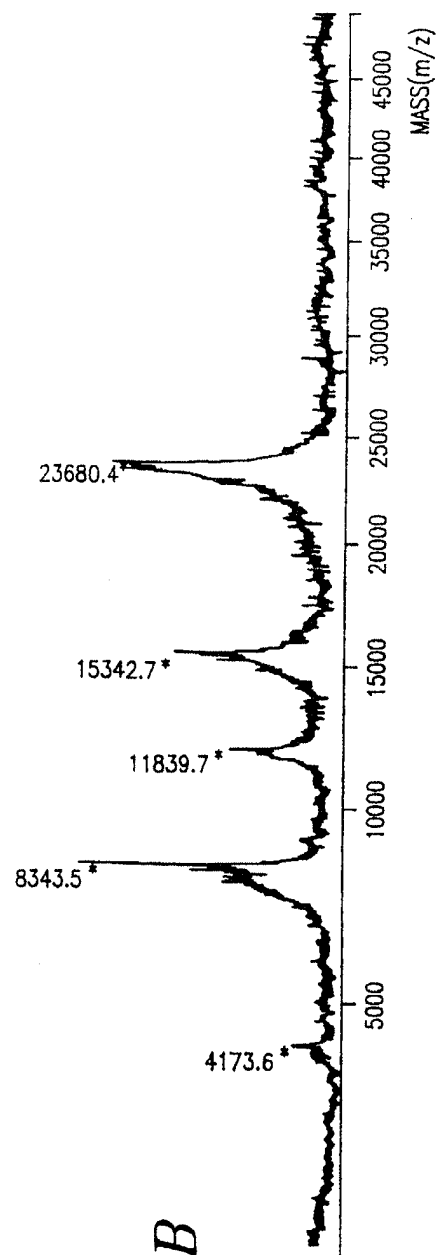


FIG. 96B

119/123

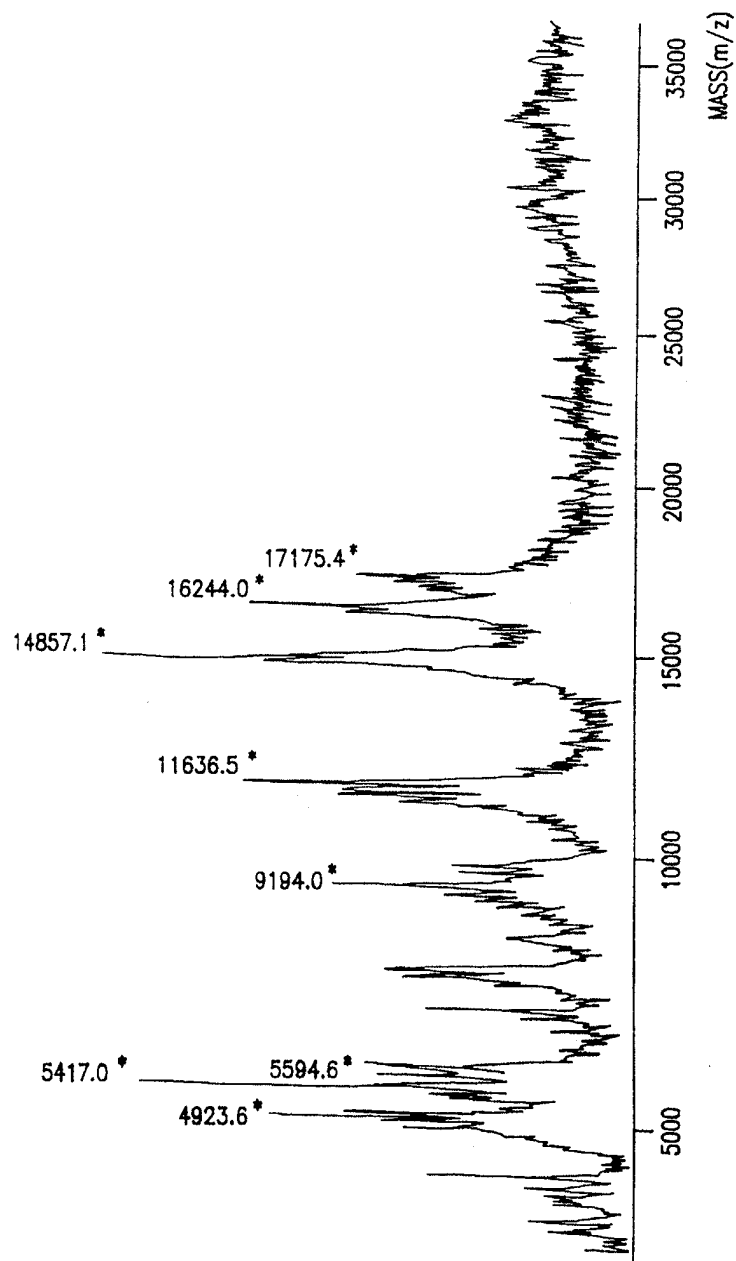


FIG. 97A

120/123

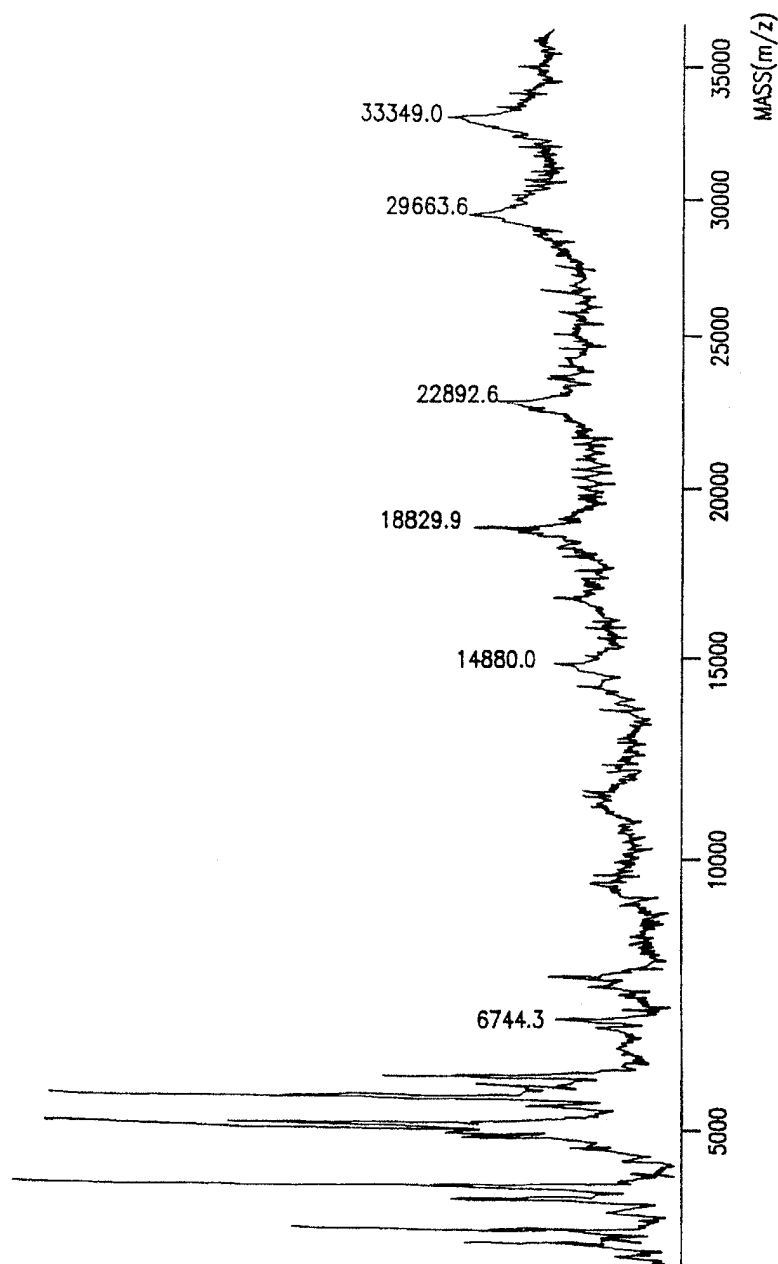


FIG. 97B

121/ 123

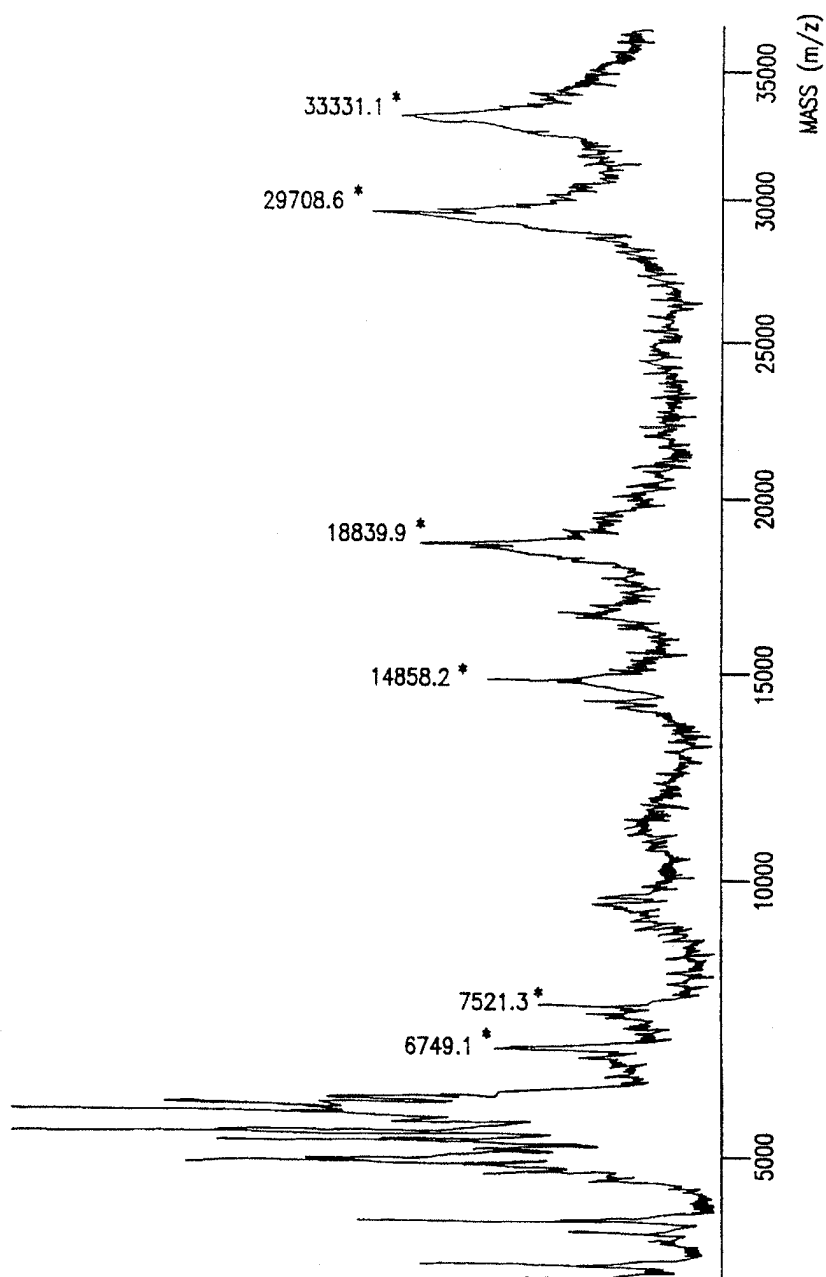


FIG. 98

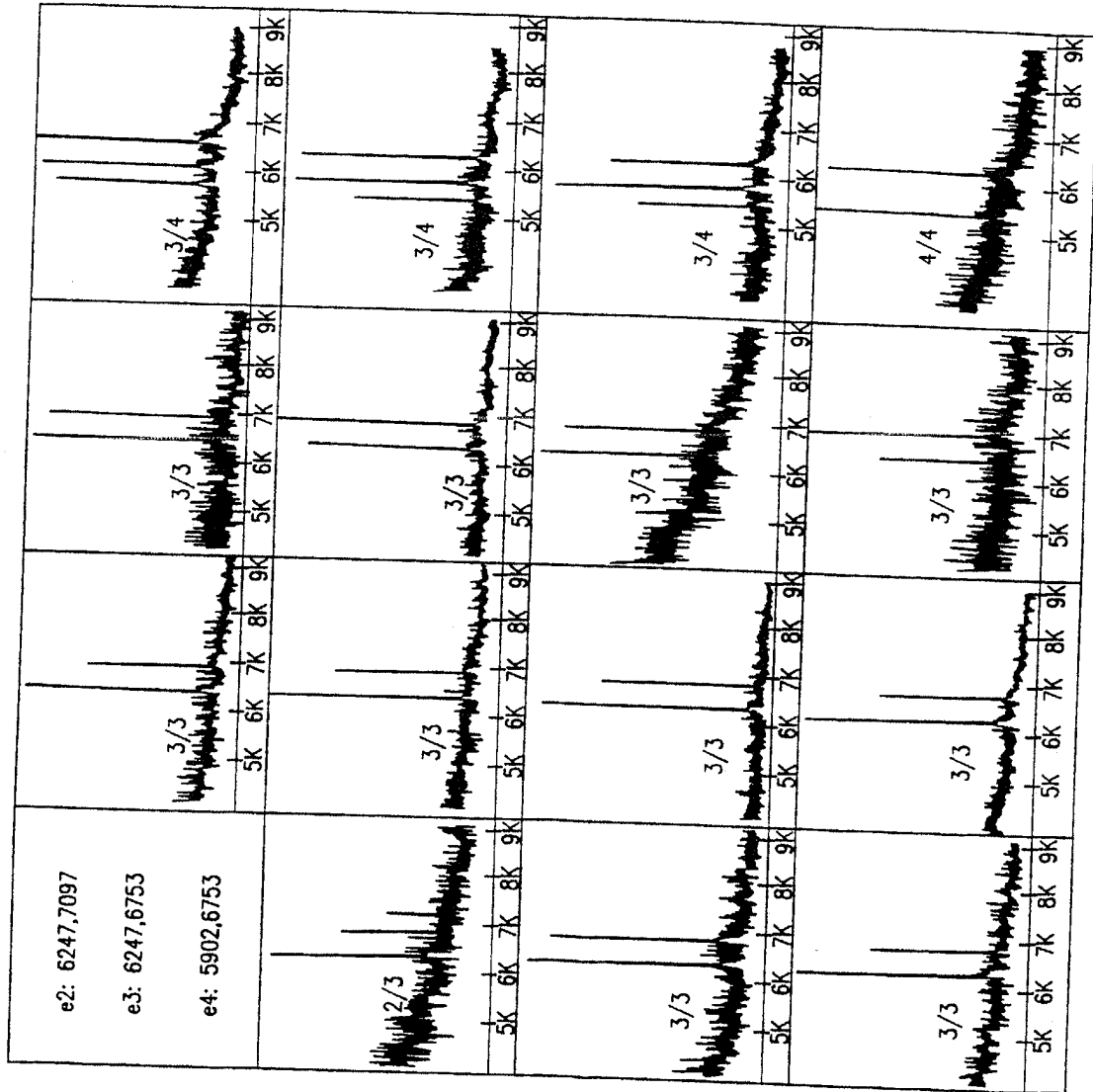


FIG. 99

Apo E
CODONS 112&158
• ISOTHERMAL PROBE
• PIN TRANSFER (~10nL)
• MALDI-MS

70% EXPECTED
ALLELIC FREQUENCY

123/123

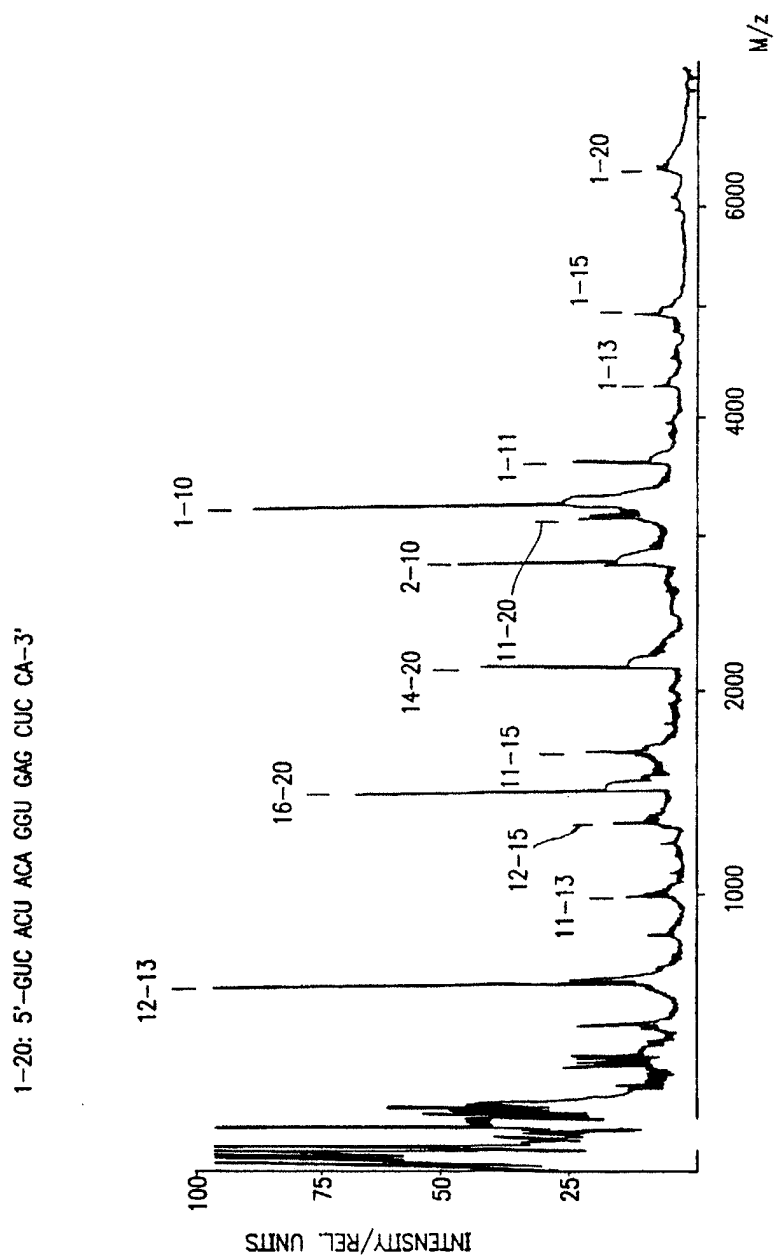


FIG. 100

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68, C07H 21/00, C07F 9/24		A3	(11) International Publication Number: WO 98/20166 (43) International Publication Date: 14 May 1998 (14.05.98)																							
(21) International Application Number: PCT/US97/20444 (22) International Filing Date: 6 November 1997 (06.11.97)			[AT/US]; 3899 Haines Street #8-308, San Diego, CA 92109 (US). JURINKE, Christian [DE/DE]; Grope Hall 68, D-22115 Hamburg (DE). VAN DEN BOOM, Dirk [DE/DE]; Forsthausstrasse 8, D-63303 Preiech (DE). XIANG, Guobing [CN/US]; Apartment: 23, 11381 Zapata Avenue, San Diego, CA 92126 (US). LOUGH, David, M. [GB/GB]; 32 Deanhead Road, Eyemouth, Berwickshire TD14 55A (GB).																							
(30) Priority Data: <table border="0"><tr><td>08/744,481</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,036</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,055</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/744,590</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/786,988</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/787,639</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/933,792</td><td>19 September 1997 (19.09.97)</td><td>US</td></tr><tr><td>08/947,801</td><td>8 October 1997 (08.10.97)</td><td>US</td></tr></table>				08/744,481	6 November 1996 (06.11.96)	US	08/746,036	6 November 1996 (06.11.96)	US	08/746,055	6 November 1996 (06.11.96)	US	08/744,590	6 November 1996 (06.11.96)	US	08/786,988	23 January 1997 (23.01.97)	US	08/787,639	23 January 1997 (23.01.97)	US	08/933,792	19 September 1997 (19.09.97)	US	08/947,801	8 October 1997 (08.10.97)
08/744,481	6 November 1996 (06.11.96)	US																								
08/746,036	6 November 1996 (06.11.96)	US																								
08/746,055	6 November 1996 (06.11.96)	US																								
08/744,590	6 November 1996 (06.11.96)	US																								
08/786,988	23 January 1997 (23.01.97)	US																								
08/787,639	23 January 1997 (23.01.97)	US																								
08/933,792	19 September 1997 (19.09.97)	US																								
08/947,801	8 October 1997 (08.10.97)	US																								
(71) Applicant (for all designated States except US): SEQUENOM, INC. [US/US]; 11555 Sorrento Valley Road, San Diego, CA 92121 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): KOSTER, Hubert [DE/US]; 8636 C Via Mallorca Drive, La Jolla, CA 92037 (US). TANG, Kai [CN/US]; 8521 Summerdale Road #241, San Diego, CA 92126 (US). FU, Dong-Jing [CN/US]; 10615 Dabney Drive #21, San Diego, CA 92126 (US). SIEGERT, Carston, W. [DE/US]; Geielstrasse 42, D-22303 Hamburg (DE). LITTLE, Daniel, P. [US/US]; 393 Glendale Lake Road, Patton, PA 18668 (US). HIGGINS, G., Scott [GB/DE]; Haselweg 1, D-22880 Weidel (DE). BRAUN, Andreas [DE/US]; 13232 Benchley Road, San Diego, CA 92130 (US). DAMHOFFER-DEMAR, Brigitte			(74) Agent: SEIDMAN, Stephanie, L.; Brown Martin Haller & McClain, 1660 Union Street, San Diego, CA 92101-2926 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).																							
			Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>																							
			(88) Date of publication of the international search report: 22 October 1998 (22.10.98)																							
(54) Title: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY																										
(57) Abstract <p>Fast and highly accurate mass spectrometry-based processes for detecting a particular nucleic acid sequence in a biological sample are provided. Depending on the sequence to be detected, the processes can be used, for example, to diagnose a genetic disease or chromosomal abnormality; a predisposition to a disease or condition, infection by a pathogenic organism, or for determining identity or heredity.</p>																										

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NI	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20444

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/68 C07H21/00 C07F9/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

29 July 1998

Date of mailing of the international search report

28.08.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Osborne, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/20444

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 29431 A (SEQUENOM INC) 26 September 1996	1
X	pages 25-54	2-7,
	see page 9, line 33 - page 10, line 2;	11-18,
	claims 1-49; figure 9	82,83
X	see page 15, line 34 - page 18, line 10;	19-34,
	examples 5,8	82,83
X	see example 8	42
X	see page 16, line 4; figures 6A,8	47
X	page 36, ln 33	48,49,
	see page 26, line 7	80,81
X	see page 16 - page 18, line 10; figures	50-64,
	3-9	68-70
Y	see page 21, line 21 - line 23	35-37
Y	see example 7	38,39
Y	see claims 1-49	40,41,
		43-47,
		65-67,
		71-79

Y	WO 94 16101 A (KOESTER HUBERT) 21 July 1994	1-7,
		11-18,
		65-79,
		82,83
	see the whole document	

Y	WO 96 32504 A (UNIV BOSTON) 17 October 1996	1-7,
		11-18,
		66-79,
		82,83
	especially page 21, lns 12-24.	
	see the whole document	

Y	WO 95 13381 A (GERON CORP) 18 May 1995	35-37,
		82,83
	see page 9, line 1 - line 33	
	see page 29, line 34 - page 32, line 5	
	see claims 1-4; example 3	

Y	FENG J ET AL: "THE RNA COMPONENT OF HUMAN TELOMERASE"	35-37,
	SCIENCE,	82,83
	vol. 269, no. 5228, 1 September 1995,	
	pages 1236-1241, XP000645335	
	see page 1236, paragraph 1 - paragraph 4	

Y	SYVANEN A -C ET AL: "DETECTION OF POINT MUTATIONS BY SOLID-PHASE METHODS"	35,38,
	HUMAN MUTATION,	39,82,83
	vol. 3, no. 3, 1 January 1994, pages	
	172-179, XP000600258	
	see the whole document	

	-/--	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/20444

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PASINI B ET AL: "RET mutations in human disease" TRENDS IN GENETICS, vol. 12, no. 4, April 1996, pages 138-144, XP002072975 see page 141, paragraph 2 - page 144 ---	35,38, 39,82,83
Y	DE 44 31 174 A (DEUTSCHES KREBSFORSCH) 7 March 1996 see abstract and claim 1 ---	40,41
Y	NAITO H ET AL: "Detection of tyrosine hydroxylase mRNA and minimal Neuroblastoma cells by reverse transcription-polymerase chain reaction " EUROPEAN JOURNAL OF CANCER, vol. 27, June 1991, pages 762-65, XP002073102 see the whole document ---	40,41
Y	GB 2 260 811 A (YORKSHIRE CANCER RESEARCH CAMP ;UNIV LEEDS (GB)) 28 April 1993 see the whole document ---	40,41
Y	WO 96 17080 A (IMP CANCER RES TECH ;SELBY PETER JOHN (GB); BURCHILL SUSAN ANN (GB) 6 June 1996 see page 3, line 1 - line 2 ---	35,40, 41,82,83
Y	NELSON R ET AL: "Time-of-flight Mass spectrometry of nucleic acids by laser ablation and ionization from a frozen aqueous matrix " RAPID COMMUNICATIONS IN MASS SPECTROMETRY, vol. 4, September 1990, pages 348-351, XP002072976 see abstract ---	42,82,83
Y	TANG K ET AL: "MATRIX-ASSISTED LASER DESORPTION/IONIZATION OF RESTRICTION ENZYME-DIGESTED DNA" RAPID COMMUNICATIONS IN MASS SPECTROMETRY, vol. 8, no. 2, February 1994, pages 183-186, XP000608266 see the whole document ---	42,82,83
Y	SIEGERT C ET AL: "Matrix-assisted laser desorption/ionization Time-of-flight Mass Spectrometry for the detection of polymerase chain reaction products containing 7-Deazapurine moieties" ANALYTICAL BIOCHEMISTRY, vol. 243, 1996, pages 55-65, XP002072977 see the whole document ---	42,82,83
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/20444

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95 15400 A (UNIV JOHNS HOPKINS) 8 June 1995 see abstract and claims 1-19 ---	43-45, 82,83
Y	WO 96 10648 A (PROMEGA CORP) 11 April 1996 see claims 1,23 ---	43-45, 82,83
Y	WO 93 23563 A (CEMU BIOTEKNIK AB ;UHLEN MATHIAS (SE); PETTERSSON BERTIL (SE)) 25 November 1993 see claims 1-7; figure 1 ---	46,82,83
Y	DE 44 38 630 A (PACHMANN KATHARINA DR ;GOEHLY URSULA (DE)) 2 May 1996 see claim 1; figures 1A,1B ---	46,82,83
Y	EP 0 593 789 A (SUMITOMO METAL IND) 27 April 1994 see abstract, claims 1 and 2 ---	46,82,83
Y	WO 96 15262 A (MEDINNOVA SF ;DZIEGLEWSKA HANNA EVA (GB); BREIVIK JARLE (NO); GAUD) 23 May 1996 see page 12, paragraph 4 - page 13, paragraph 2 ---	46,82,83
Y	WO 89 06700 A (GENENTECH INC) 27 July 1989 see the whole document ---	47,82,83
Y	WO 89 03432 A (US ENERGY) 20 April 1989 see claims 1-21; figures 1,2 ---	65,68-70
Y	US 5 288 644 A (BEAVIS RONALD C ET AL) 22 February 1994 see claims 1-8 ---	66,67
X	WO 94 21822 A (KOESTER HUBERT) 29 September 1994 see claims 1-55 ---	65-79
X	US 5 430 136 A (URDEA MICHAEL S ET AL) 4 July 1995 see column 4, line 28 - column 9, line 52 ---	84-87
X	WO 95 31429 A (UNIV BOSTON) 23 November 1995 see claims 1,19 ---	84-87
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 97/20444

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ORDOUKHANIAN P ET AL: "Design and synthesis of a versatile photocleavable DNA building block. application to phototriggered hybridization" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 117, 1995, pages 9570-71, XP002072978 see the whole document ---	84-87
E	WO 97 42348 A (SEQUENOM INC) 13 November 1997 see claims 1-48 -----	1,7-10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 97/ 20444

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

SEE ANNEXES

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☒ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

It should be further noted that "claims 82-83" as designated herein refer to two claims 82 and two claims 83 which were filed as follows, claims 82, 83 followed by a second claim 82 and a second claim 83)

1. Claims 1-18, partially 82-83:

A method for determining the sequence of a target nucleic acid involving the generation of base specifically terminated fragments.

2. Claims 19-34, partially 82-83:

A method for detecting a target nucleic acid present in a biological sample based on a nested polymerase chain amplification reaction.

3. Claim 35 partially (in that it relates to the detection of neoplasia/malignancies by detecting telomerase), claims 36 and 37, and partially 82-83:

An assay for the detection of neoplasia/malignancies based on telomerase specific extension of a substrate primer and a subsequent amplification of the telomerase specific extension product by PCR.

4. Claim 35 partially (in that it relates to the detection of neoplasia/malignancies by detecting mutation of a proto-oncogene), claims 38 and 39, and partially claims 82-83:

An assay for the detection of neoplasia involving mutation analysis of mutant or wild-type alleles by primer extension reaction by a Sanger type sequencing protocol.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

5. Claim 35 partially (in that it relates to the detection of neoplasia/malignancies by detecting expression of a tumour-specific gene in a specific tissue type), claims 40 and 41, and partially claims 82-83:

An amplification based assay for the expression of the tyrosine hydroxylase gene in bone marrow cells as indicative of a neuroblastoma.

6. Claim 42, partially claims 82-83:

A method for directly detecting double stranded nucleic acid using Maldi-TOF mass spectrometry.

7. Claims 43-45, partially claims 82-83:

A method for comparing DNA relatedness by amplification of microsatellite DNA repeat sequences.

8. Claim 46, partially claims 82-83:

A method for detecting mutations based on target amplification using a primer that introduces a unique endonuclease restriction site into amplified target and a combination of a Sanger sequencing protocol and endonuclease digestion.

9. Claim 47, partially claims 82-83:

A method for the amplification and detection of a nucleic acid based on the synthesis of RNA using a primer containing a RNA polymerase promoter sequence.

10. Claims 48, 49, 80 and 81, partially 82-83:

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Primers per se for mass spectrometry comprising a mass modifying moiety.

11. Claims 50-64, partially 68-70, partially 73-79, partially claims 82-83:

Method for detecting a target nucleic acid sequence involving hybridisation to a detector oligonucleotide.

12. Claims 65-67, partially 68-70, 71-72, partially 73-79, partially claims 82-83:

Methods for determining a nucleic acid sequence involving exonuclease digestion.

13. Claims 84-94:

Photolabile linkers per se for use in immobilisation of nucleic acids to solid supports.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/20444

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9629431 A	26-09-1996	US 5605798 A AU 5365196 A CA 2214359 A EP 0815261 A	25-02-1997 08-10-1996 26-09-1996 07-01-1998
WO 9416101 A	21-07-1994	AU 5992994 A CA 2153387 A EP 0679196 A JP 8509857 T US 5547835 A US 5605798 A US 5691141 A	15-08-1994 21-07-1994 02-11-1995 22-10-1996 20-08-1996 25-02-1997 25-11-1997
WO 9632504 A	17-10-1996	AU 5544696 A EP 0830460 A	30-10-1996 25-03-1998
WO 9513381 A	18-05-1995	US 5645986 A US 5629154 A AU 1178195 A AU 682082 B AU 1209095 A AU 1330795 A AU 6058298 A CA 2173872 A EP 0728207 A JP 9502102 T WO 9513382 A US 5648215 A US 5686306 A US 5639613 A US 5693474 A WO 9513383 A	08-07-1997 13-05-1997 29-05-1995 18-09-1997 29-05-1995 29-05-1995 04-06-1998 18-05-1995 28-08-1996 04-03-1997 18-05-1995 15-07-1997 11-11-1997 17-06-1997 02-12-1997 18-05-1995
DE 4431174 A	07-03-1996	NONE	
GB 2260811 A	28-04-1993	NONE	
WO 9617080 A	06-06-1996	NONE	
WO 9515400 A	08-06-1995	NONE	

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 97/20444

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9610648 A	11-04-1996	AU 3998195 A CA 2118048 A	26-04-1996 31-03-1996
WO 9323563 A	25-11-1993	AU 4068293 A CA 2135606 A EP 0641391 A JP 8500725 T	13-12-1993 25-11-1993 08-03-1995 30-01-1996
DE 4438630 A	02-05-1996	NONE	
EP 0593789 A	27-04-1994	JP 5308999 A WO 9323567 A	22-11-1993 25-11-1993
WO 9615262 A	23-05-1996	AU 3851495 A CA 2205017 A EP 0791074 A	06-06-1996 23-05-1996 27-08-1997
WO 8906700 A	27-07-1989	AU 3058589 A DE 68908054 T DK 463089 A EP 0359789 A JP 2006724 A JP 2503054 T NO 300782 B	11-08-1989 10-03-1994 20-09-1989 28-03-1990 10-01-1990 27-09-1990 21-07-1997
WO 8903432 A	20-04-1989	US 4962037 A CA 1314247 A DE 3854743 D DE 3854743 T EP 0381693 A JP 3502041 T	09-10-1990 09-03-1993 11-01-1996 09-05-1996 16-08-1990 16-05-1991
US 5288644 A	22-02-1994	US 5453247 A	26-09-1995
WO 9421822 A	29-09-1994	AU 687801 B AU 6411694 A CA 2158642 A EP 0689610 A JP 8507926 T US 5622824 A	05-03-1998 11-10-1994 29-09-1994 03-01-1996 27-08-1996 22-04-1997

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/20444

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5430136 A	04-07-1995	US 5258506 A	02-11-1993
		US 5118605 A	02-06-1992
		US 4775619 A	04-10-1988
		CA 2088257 A	28-01-1992
		EP 0543889 A	02-06-1993
		JP 8311091 A	26-11-1996
		JP 9031090 A	04-02-1997
		JP 2552048 B	06-11-1996
		PL 170146 B	31-10-1996
		PT 98488 A,B	29-05-1992
		WO 9202528 A	20-02-1992
		US 5545730 A	13-08-1996
		US 5578717 A	26-11-1996
		US 5552538 A	03-09-1996
		US 5367066 A	22-11-1994
		AT 133714 T	15-02-1996
		DE 3854969 D	14-03-1996
		DE 3854969 T	30-05-1996
		EP 0360940 A	04-04-1990
		EP 0703296 A	27-03-1996
		ES 2083955 T	01-05-1996
		JP 2092300 A	03-04-1990
		JP 2676535 B	17-11-1997
		US 5380833 A	10-01-1995
WO 9531429 A	23-11-1995	US 5643722 A	01-07-1997
		AU 2635995 A	05-12-1995
		CA 2189848 A	23-11-1995
		EP 0763009 A	19-03-1997
		JP 10500409 T	13-01-1998
WO 9742348 A	13-11-1997	AU 3003497 A	26-11-1997